



**KÜMELEME ANALİZİ VE SAĞLIK ALANINDA BİR
UYGULAMA**

Tuba USLU

Biyoistatistik ve Tıp Bilişimi Anabilim Dalı

**Tez Danışmanı
Dr. Öğr. Üyesi Harika Güzde GÖZÜKARA BAĞ**

Yüksek Lisans Tezi -2019

T.C.
İNÖNÜ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ

KÜMELEME ANALİZİ VE SAĞLIK ALANINDA BİR UYGULAMA

Tuba USLU

Biyostatistik ve Tıp Bilişimi Anabilim Dalı

Yüksek Lisans Tezi

Tez Danışmanı

Dr. Öğr. Üyesi Harika Gözde GÖZÜKARA BAĞ

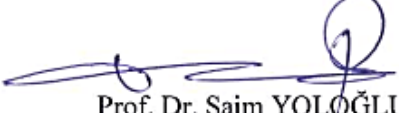
MALATYA

2019

KABUL VE ONAY SAYFASI

İnönü Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik ve Tıp Bilişimi Anabilim Dalı Yüksek Lisans Programı çerçevesinde yürütülmüş olan; **Tuba USLU'nun "Kümeleme Analizi ve Sağlık Alanında Bir Uygulama"** konulu bu çalışması, aşağıdaki jüri tarafından Yüksek Lisans tezi olarak kabul edilmiştir.

Tez Savunma Tarihi: 20/06/2019


Prof. Dr. Saim YOLOĞLU
İnönü Üniversitesi
Jüri Başkanı


Dr. Öğr. Üyesi Mehmet Onur KAYA
Fırat Üniversitesi
Üye


Dr. Öğr. Üyesi Harika G. GÖZÜKARA BAĞ
İnönü Üniversitesi
Tez Danışmanı
Üye

ONAY

Bu tez, İnönü Üniversitesi Lisansüstü Eğitim-Öğretim Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından kabul edilmiş ve Enstitü Yönetim Kurulu'nun/...../2019 tarih ve 2019/..... sayılı Kararıyla da uygun görülmüştür.

Prof. Dr. Yusuf TÜRKÖZ
Enstitü Müdürü

İÇİNDEKİLER

ÖZET	vi
ABSTRACT.....	vii
SİMGELER VE KISALTMALAR DİZİNİ	viii
ŞEKİLLER DİZİNİ	ix
TABLolar DİZİNİ.....	x
1. GİRİŞ.....	1
2. GENEL BİLGİLER	4
2.1. Uzaklık Ölçüleri.....	4
2.1.1. Öklid Uzaklığı ve Kare Öklid Uzaklığı	4
2.1.2. Minkowski Uzaklığı	5
2.1.3. Manhattan Uzaklığı.....	5
2.1.4. Mahalanobis Uzaklığı.....	5
2.1.5. Hotelling T^2 Uzaklığı.....	6
2.1.6. Chebychev Uzaklığı.....	6
2.1.7. Canberra Uzaklığı	6
2.1.8. Pearson Uzaklığı.....	7
2.1.9. Karesel Pearson Uzaklığı.....	7
2.1.10. Czekanowski Katsayısı	7
2.2. Küme Sayısının Belirlenmesi	7
2.3. Hiyerarşik Kümeleme Analizi	8
2.3.1. Tek Bağlantı Yöntemi (Nearest Neighbor Method)	9
2.3.2. Tam Bağlantı Yöntemi (Furthest Neighbor Method)	9
2.3.3. Ortalama Bağlantı Yöntemi (Average Linkage Method)	10
2.3.4. Medyan Bağlantı Kümeleme Yöntemi (Median Linkage)	10

2.3.5. Küme Merkezleri Yöntemi (Centroid Clustering).....	11
2.4. Hiyerarşik Olmayan Kümeleme Analizi.....	11
2.4.1. En Çok Olabilirlik Tekniği	11
3. MATERYAL VE METOT	12
3.1. Ward's Yöntemi	12
3.2. K-Ortalama Yöntemi	12
3.3. İki Aşamalı Kümeleme Analizi	13
3.4. Veri Analizi.....	17
4. BULGULAR.....	20
5. TARTIŞMA	40
6. SONUÇ VE ÖNERİLER.....	41
KAYNAKLAR	43
EKLER.....	47
EK-1. Özgeçmiş.....	47
EK-2. Etik Kurul Almama Gerekçesi	48

TEŐEKKÜR

Bu alıőmanın planlanması ve yürütülmesi sürecinde benden destek ve ilgilerini esirgemeyen, lisansüstü eğitimim boyunca yardımını ve bilgi birikimini esirgemeyen, bilgi ve tecrübeleriyle beni yönlendiren değerli hocam Dr. Öğr. Üyesi Harika Gözde GÖZÜKARA BAĞ'a teşekkür ederim.

Lisansüstü eğitimim boyunca göstermiş oldukları yakın ilgi ve vermiş oldukları destek ve emeklerden dolayı Prof. Dr. Saim YOLOĞLU ve Prof. Dr. Cemil ÇOLAK hocalarıma teşekkürlerimi sunuyorum.

Ögr. Gör. Tuba USLU

Malatya-2019

ÖZET

Kümeleme analizi ve sağlık alanında bir uygulama

Amaç: Bu çalışmanın amacı, kümeleme analizi yöntemlerini tanıtmak ve ülkelerin sağlık göstergelerine göre kümeleme analizi yöntemlerinin uygulamasını göstermektir. Böylece verilere uygulanan kümeleme analizi yöntemleri sonuçları ve doğru analiz yöntemi ile geçerli bulgular elde edilebileceği örneklenecektir.

Materyal ve Metot: Bu çalışmada, ülkeler sağlık göstergesi olarak kabul edilen değişkenlere göre kümelenecek olup veri kaynağı olarak Dünya Sağlık Örgütü ve Dünya Bankası web sitelerinde yayınlanmış en güncel veriler kullanılmıştır. Çalışmada Ward's yöntemi, k-ortalama yöntemi ve iki aşamalı kümeleme yöntemi kullanılmıştır.

Bulgular: Ward's yöntemi uygulanmadan önce veriler standartlaştırılmıştır ve ardından yöntem uygulanmıştır. Yöntem sonucunda iki küme oluşmuştur. Daha sonra k-ortalama ve iki aşamalı yöntem uygulanmış olup yine iki küme oluştuğu gözlenmiştir. Oluşan kümelerdeki ülkelerin sağlık göstergelerinin tamamının önemli düzeyde etkin olduğu ANOVA sonuçları ile görülmüştür ($p < 0.05$).

Sonuç: Hiyerarşik kümeleme yöntemleri uygulanırken verileri standartlaştırma işleminin yapılıp yapılmayacağı incelenmeli ve bu durumu giderebilecek yöntem kullanılmalıdır. Aksi halde kümeleme yöntemi sonucunda oluşacak kümeler yanlış sonuçlara götürebilecektir. Ayrıca gözlem sayısına, değişken sayısına ve değişkenlerin veri tipine bağlı olarak hangi yöntemin uygulanacağı belirlenmelidir. Üç kümeleme yönteminin uygulandığı bu çalışmanın sonucunda, tüm yöntemler için aynı kümelene yapısı elde edilmiştir.

Anahtar Kelimeler: Kümeleme Analizi, Uzaklık Ölçüsü, Wards Metodu, k-ortalama metodu

ABSTRACT

Cluster Analysis and an application in healthcare

Aim: The aim of this study is to introduce cluster analysis methods and to show the application of cluster analysis methods according to health indicators of countries. Thus, the correct analysis method can be obtained by the results of the cluster analysis methods applied to the data.

Material and Method: In this study, the countries will be clustered according to the variables considered as health indicators and the most recent data published on the World Health Organization and World Bank websites are used as data source. Ward's method, k-means method and two-step clustering method were used in the study.

Results: Before Ward's method was applied, the data were standardized and then the method was applied. As a result of the method, two clusters were formed. Then, k-means and two-step method was applied and two clusters were observed. ANOVA results showed that all of the health indicators of the countries in the clusters formed were significantly effective ($p < 0.05$).

Conclusion: When applying hierarchical clustering methods, the standardization process of the data should be examined and the method that can eliminate this situation should be used. Otherwise, the clustering method will lead to incorrect results. Also, depending on the number of observations, variables and the data type, it should be determined which method should be applied. Based on the results of this study in which three clustering methods applied, all methods resulted with the same cluster structure.

Keywords: Cluster Analysis, Distance Measure, Ward's Method, k-means method

SİMGELER VE KISALTMALAR DİZİNİ

ANOVA	: Varyans Analizi
DB	: Dünya Bankası
DSÖ	: Dünya Sağlık Örgütü



ŞEKİLLER DİZİNİ

Şekil No	Sayfa No
Şekil 4.1. Ülkelerin Ağaç Grafiği (Dendrogram)	26
Şekil 4.2. Model Özeti ve Kümeleme Kalitesi	35
Şekil 4.3. Oluşan Küme Büyüklükleri	37
Şekil 4.4. İki Aşamalı Yönteme Göre Oluşan Küme Ortalamaları ve Sağlık Göstergesi Değişkenlerinin Önem Dereceleri.....	38
Şekil 4.5. Kümelerdeki Sağlık Göstergesi Değişkenlerinin Önem Derecelerinin Grafikleri	39

TABLULAR DİZİNİ

Tablo No	Sayfa No
Tablo 3.1. Ülkelere Göre Sağlık Göstergeleri Veri Seti	18
Tablo 4.1. Sağlık Göstergelerinin Tanımlayıcı İstatistikleri.....	20
Tablo 4.2. Standardize Edilen Veriler	21
Tablo 4.3. Standardize Edilmiş Verilere Göre Ülkelerin Kare Öklit Uzaklık Matrisi ...	24
Tablo 4.4. Ward's Yöntemi Sonucunda Kümelerde Yer Alan Ülkeler.....	27
Tablo 4.5. Kümelerin Oluşması Aglomeratif Tablo (Birleştirme Tablosu)	28
Tablo 4.6. K-Ortalama Yöntemi Sonucunda Elde Edilen Kümeler ve Uzaklıkları.....	30
Tablo 4.7. K-Ortalama Yöntemi Sonucunda Kümelerde Yer Alan Ülkeler.....	31
Tablo 4.8. Son Küme Merkezleri.....	31
Tablo 4.9. Son Küme Merkezleri Arasındaki Uzaklıklar	32
Tablo 4.10. ANOVA tablosu	33
Tablo 4.11. İki Aşamalı Kümeleme Yöntemi Sonucu Oluşan BIC Kriter Tablosu	34
Tablo 4.12. Küme Merkezleri (Centroids).....	35
Tablo 4.13. İki Aşamalı Kümeleme Analizi Sonucunda Oluşan Kümeler	39

1. GİRİŞ

Kümeleme analizi uygulamada birçok alanda karşımıza çıkmaktadır. Kümeleme analizi ilk olarak biyoloji ve ekoloji alanında, daha sonra ise tıp, antropoloji, jeoloji, çevre, gıda bilimi ve mühendislik gibi birçok alanda kullanılmıştır (1,2). Sağlık alanında karşılaştığımız durumlar; illeri sağlık verilerine göre belirli sayıda kümelerde sınıflandırarak en kötü durumda olan illeri tespit etmede, tıbbi teşhislerde, biyolojik veri analizinde, medikal görüntü segmentasyonunda, hasta veri tabanı yöntemi ve hastane kaynak yönetimi gibi çalışmalar kümeleme analizine örnek gösterilebilir (3-6). Ayrıca, tıbbi araştırmalarda çeşitli hastalıklarda fenotip ile grupları tanımlamak için de yine kümeleme analizi kullanılmıştır (7,8).

Kümeleme analizi, nesnelerin veya gözlemlerin benzerlik temelinde homojen gruplara ayrıldığı istatistiksel veri bölümlenme yöntemlerini ifade eder ve çok değişkenli bir veri kümesinden (gözlemler, nesneler) küme grupları oluşturmak için kullanılan çok değişkenli bir yöntemdir (1,9,10). Bu nedenle kümeleme analizi, doğal olarak benzerlik kavramına bağlıdır (11). Her gruptaki gözlemler birbirine benzerdir, ancak kümelerin kendileri birbirinden farklıdır. Verileri benzerliğe göre gruplara ayırmak için birçok algoritma vardır ve bunlar önemli ölçüde farklıdır (1). Kümeleme analizi, bir tür veri azaltma tekniğidir (2).

Kümeleme analizinin genel amacı; toplanan çok sayıdaki gruplanmamış gözlemlerden oluşan veriyi gözlemlerin benzerliklerine göre anlamlı gruplardan oluşan özel alt kümelere bölerek veriyi indirgemektir (12). Böylece araştırmacı en az bilgi kaybıyla, daha net ve anlaşılabilir tanımlı gözlemlere sahip olacaktır. Kümeleme analizi ilk kez 1939 yılında Tryon tarafından kullanılmıştır. Kümeleme analizine genel ilgi 1960'lerde artmıştır ve bu da analiz olanaklarını genişleten birkaç yeni algoritmanın geliştirilmesine yol açmıştır (2). 1963 yılında Robert Sokal ve Peter Sneath'ın yazdığı "Sayısal Sınıflandırma İlminin Temelleri" adlı kitap bu alanda önemli bir adım olmuştur (12).

Kümeleme teknikleri, geniş bir veri setini gruplara ayırmanın gerekliliği olan çok çeşitli araştırma programlarına uygulanmıştır (1). Grupların sayısı ve her kümedeki gözlem sayısı gruplama süreci başlamadan önce bilinmemektedir. Küme analizi, doğru grupları sınıflandırmak ve tanımlamak için değerlidir (1).

Kümeleme analizi ile grupların yani kümelerin içsel olarak homojen (bir kümenin içinde), ancak dışsal olarak farklı (kümeler arasında) olduğu gruplar oluşturulmaktadır (13). Kümeleme analizi, gözlemlenen verileri iki veya daha fazla homojen gruba ayırmaya yardımcı olan çok kullanışlı bir yöntemdir ve bu analizin avantajı; hangi nesnenin hangi kümelere ait olduğu konusunda önceden bilgi gerektirmemesidir (14).

Kümeleme analizi, sınıflandırma analizinden farklıdır. Çünkü sınıflandırma analizinde grupların sayısı bilinir ve hedef her bir gözlemi bir önceden tanımlanmış gruba yeniden atamaktır. Kümeleme analizinde ise önceden tanımlanmış grupların varsayımı yoktur ve grupların sayısı, gözlemler arasındaki benzerliğe dayanarak belirlenmiştir (1).

Kümeleme analizinde diğer çok değişkenli istatistik analizlerde olduğu gibi verilerin normalliği varsayımı fazla önemli olmayıp uzaklık değerlerinin normalliği yeterli görülmektedir (15). Analizde kullanılan değişkenler arasındaki ilişkilerin doğrusal olma zorunluluğu yoktur ve analiz seçilen kümeleme yöntemine göre sınıflayıcı (nominal), sıralayıcı (ordinal), aralıklı (interval), oransal (ratio) veya kategorik ölçekle ölçülen verilere uygulanabilmektedir (16).

Kümeleme analizi, değişkenler arasındaki bağımlılıkların incelenmesine, benzer varlıkların küme olarak adlandırılan daha homojen gruplara dâhil edilmesine dayanmaktadır. Süreç ise (17);

1. Analizin amacının tanımı, test edilecek varsayımın ve en önemli değişkenlerin seçilmesi

2. Bir standardizasyon yöntemi uygulayarak sayısal değerleri işleme: Örneklemdeki gözlemler arasındaki benzerlik ilişkilerinin yanlış temsil edilme riskini azaltmak için değişkenler farklı ölçü birimlerinde ifade edildiğinde standardizasyon uygulanır. Bu nedenle, değişkenler boyutsuz hale gelecektir. Standardizasyon prosedürünün bir başka avantajı; yanlış sonuçların üretilmesine yatkın olan aşırı değerlerin ortadan kaldırılmasıyla, değişkenlerin etkisinin eşitlenmesidir. Standardizasyon başarısız olursa, bir değişkenin değerleri diğerlerinden büyük bir aralıkta değişirse, o zaman bu değişkenler gözlemler arasındaki benzerliklerin kurulmasında, sonuçların denatüre edilmesinde daha büyük bir öneme sahip olacaktır.

3. Bir kümeleme yönteminin seçilmesi: Yöntem belirlenirken yapılması gerekenler;

- K-ortalama kümeleme (hiyerarşik olmayan kümeleme) yöntemi küme sayısının önceden belirlenmesini gerektirir. Gözlem sayısı 1000'i aştığında tavsiye edilir.
- Hiyerarşik kümeleme yöntemi, gözlemleri hiyerarşik bir yapıya göre gruplar.
- Büyük veri kümeleri veya metin değişkenleri için çoğunlukla iki aşamalı kümeleme yöntemi uygulanır.

4. Veri toplamak için uygun bir yöntem seçilmesi: Tek bağlantı, tam bağlantı ve ortalama bağlantı yöntemi sık kullanılan yöntemlerdir.

5. Veri tipine göre (aralıklı, sayısal, ikili değişkenler), gözlemler arasındaki uzaklık/benzerlik için bir ölçü birimi veya algoritma seçimi: Kişi, benzerlik katsayısı veya benzerlik açısından anlamını yorumlayabileceği bir benzerlik katsayısını hesaplamalıdır.

6. Dendogramın yorumlanması ve optimal sayıda kümenin belirlenmesi: Doğru sayıda kümenin oluşturulması, deneyimine bağlı olarak öznel bir süreçtir.

Kümeleme analizi, seçilen veri kümesi içindeki yapıları tanımlamaya çalışan çok kullanışlı bir yöntemdir. Benzer türdeki nesnelere gruplandırmak için farklı algoritmalar ve yöntemler kullanılır. Bu yöntemlere genellikle seçilen verileri anlamlı yapılara ayıran bir segmentasyon analizi denir. Bu yöntemler bağımsız veya bağımlı değişkenler arasında herhangi bir ayırım yapmazlar (14).

Kümeleme analizi, farklı nesnelere sıralamaya çalışır. Kümeleme analizi, nesnelere homojen grupları bir şekilde aynı kümeleme içinde birbirine benzeyen ve diğer kümelerdeki nesnelere benzemeyen bir şekilde tanımlar (14).

2.GENEL BİLGİLER

Kümeleme analizi için birçok algoritma öne sürülmüştür. Ancak alan yazında bu algoritmalar iki ana başlık altında toplanmıştır (18). Bunlardan ilki, dendogram oluşturan hiyerarşik kümeleme yöntemleri, diğeri ise hiyerarşik olmayan kümeleme yöntemleri olarak adlandırılmaktadır. Her iki yöntemde de ortak amaç kümeler arasındaki farklılıkları ve kümeler içi benzerlikleri en yüksek düzeye çıkarmaktır. Diğeri bir ifadeyle, küme içi homojenlik artırılırken kümeler arası homojenlik ise azaltılmaktadır. Hangi yöntemin kullanılacağı küme sayısına bağlı olmakla birlikte her iki yöntemin beraber kullanılması çok daha yararlıdır. Böylece, hem sonuçları hem de iki yöntemden hangisinin daha uygun sonuçlar verdiğini karşılaştırmak mümkün olmaktadır (19).

2.1. Uzaklık Ölçüleri

Bir araştırmacının yapması gereken ilk adım, durumlar arasındaki mesafeyi veya benzerliği hesaplamak için kullanılacak istatistiği belirlemektir. İki durum arasındaki mesafe azaldıkça, bunların benzerlikleri sırasıyla artmalıdır(11). Kümeleme analizi, bir gruptaki benzer gözlemleri gruplandırmaktadır. Bu da, birbirine yakın gözlemlerin bir grup oluşturduğu anlamına gelir. Uzaklık farklılığın bir ölçüsüdür (1).

2.1.1. Öklid Uzaklığı ve Kare Öklid Uzaklığı

Öklid uzaklık fonksiyonu genellikle gözlemler arasındaki mesafeyi ölçmek için kullanılır. Sürekli değişkenler için en sık kullanılan uzaklık ölçüsü Öklid uzaklığıdır (1,20).

$A=(x_1,x_2,\dots,x_k)$ ve $B=(y_1,y_2,\dots,y_k)$ koordinatları için A ve B noktaları arasındaki mesafe Öklid uzaklığı olarak adlandırılır ve aşağıdaki şekilde gösterilir (1).

$$d = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_k - y_k)^2} \quad (1)$$

Değişkenler genellikle farklı değişkenler arasındaki varyasyonu ve korelasyonu (veya kovaryansları) hesaba katacak şekilde standartlaştırılır ve uzaklık (2)'de verilen formül ile hesaplanır.

$$d = \sqrt{\frac{(x_1 - y_1)^2}{s_{11}} + \frac{(x_2 - y_2)^2}{s_{22}} + \dots + \frac{(x_k - y_k)^2}{s_{kk}}} \quad (2)$$

Yöntemdeki her adımda, tüm gözlem çiftleri ve kümeler arasındaki Öklid mesafesi hesaplanır ve yakınlık matrisinde gösterilir. Her adımda, en küçük karesi alınmış Öklid uzaklığıyla gözlem çiftleri veya kümeler birbirleriyle birleştirilecektir. Bu, hiyerarşik kümelenmeyi uzun bir süreç haline getirir. Çünkü her adımdan sonra yakınlık matrisinin yeni eklenen kümelenmeyi hesaba katmak üzere yeniden hesaplanması gerekir (20).

Küme başına sadece bir gözlem olduğunda Kare Öklid uzaklığı hesaplaması basittir. Bununla birlikte, küme başına birden fazla gözlem olduğunda, Kare Öklid uzaklığının en iyi nasıl hesaplanacağı konusunda ek bir karar verilmelidir. Bu bağlantı ölçüsü olarak adlandırılır ve araştırmacı iki küme arasındaki bağlantıyı en iyi nasıl hesaplayacağını belirlemelidir (20).

2.1.2. Minkowski Uzaklığı

Her biri p tane sürekli (oransal ve aralıklı ölçekli) değişken içeren x_i ve x_j gözlem çifti arasındaki uzaklık $d_{ij} = d(x_i, x_j)$ olsun (21).

$$d_\lambda(x_i, x_j) = [\sum_{k=1}^p |x_{ik} - x_{jk}|^\lambda]^{1/\lambda} ; \lambda \geq 1 \text{ için} \quad (3)$$

2.1.3. Manhattan Uzaklığı

$x=(x_1, x_2, \dots)$ ve $y=(y_1, y_2, \dots)$ noktası arasındaki uzaklık formülü (21):

$$d_1(x_i, y_j) = \sum_{k=1}^p |x_{ik} - y_{jk}| \quad (4)$$

2.1.4. Mahalanobis Uzaklığı

Her biri p tane sürekli (oransal ve aralıklı ölçekli) değişken içeren x_i ve x_j gözlem çifti arasındaki uzaklık $d_{ij} = d(x_i, x_j)$ olsun (21).

$$d(x_i, x_j) = (x_i - x_j)^T S^{-1} (x_i - x_j) \quad (5)$$

T:Matrisin Transpozu

S:p×p Kovaryans Matrisi

S⁻¹:Kovaryans Matrisinin Tersi

Bu uzaklık ölçüsü aykırı noktaları da hesaplamaktadır (22).

2.1.5. Hotelling T² Uzaklığı

İki kümenin uzaklığını hesaplarken kullanılmaktadır.

$$T^2 = \frac{n_1 n_2}{n} (\bar{x}_{ik} - \bar{x}_{jk}) S^{-1} (\bar{x}_{ik} - \bar{x}_{jk}) \quad (6)$$

n₁: 1. Kümenin gözlem sayısı

n₂: 2. Kümenin gözlem sayısı

S: Verilerin toplamının standart sapması

\bar{x}_{ik} : i. gözlemin k. değişkendeki değerlerinin vektörü

\bar{x}_{jk} : j. gözlemin k. değişkendeki değerlerinin vektörü

2.1.6. Chebychev Uzaklığı

$x=(x_1,x_2,..x_n)$ ve $y=(y_1,y_2,..,y_n)$ vektörleri arasındaki maksimum uzaklığı ifade etmektedir (21).

$$d(x_i - y_i) = \max |x_i - y_i| = \lim_{k \rightarrow \infty} (\sum_{i=1}^n |x_i - y_i|^k)^{1/k} \quad (7)$$

2.1.7. Canberra Uzaklığı

Veri matrisindeki değişkenlerin tamamının pozitif olduğu durumlarda kullanılan bir uzaklık ölçüsüdür (21).

$$d(x_i - y_i) = \sum_{i=1}^n \frac{|x_i - y_i|}{x_i + y_i} \quad (8)$$

2.1.8. Pearson Uzaklığı

İki birim arasındaki uzaklık hesaplanırken aşağıdaki formül kullanılır(21).

$$d(i, j) = \sqrt{\frac{(x_{i1}-x_{j1})^2}{s_1^2} + \frac{(x_{i2}-x_{j2})^2}{s_2^2} + \dots + \frac{(x_{ip}-x_{jp})^2}{s_p^2}} \quad (9)$$

S_p : Uzaklığın hesaplandığı değişkene ait varyans

2.1.9. Karesel Pearson Uzaklığı

$$d(i, j) = \frac{(x_{i1}-x_{j1})^2}{s_1^2} + \frac{(x_{i2}-x_{j2})^2}{s_2^2} + \dots + \frac{(x_{ip}-x_{jp})^2}{s_p^2} \quad (10)$$

2.1.10. Czekanowski Katsayısı

Negatif olmayan değişkenler için kullanılan bir uzaklık ölçüsüdür (23):

$$d(x_i, x_j) = 1 - \frac{2 \sum_{k=1}^p \min(x_{ik}, x_{jk})}{\sum_{k=1}^p x_{ik} + x_{jk}} \quad (11)$$

2.2. Küme Sayısının Belirlenmesi

İdeal olması gereken küme sayısı, küme içi uzaklıklarının minimum ve kümeler arası uzaklığın maksimum olduğu andır. Ancak küme sayısının belirlenmesinde karar araştırmacıya bırakılmıştır. Küme sayısı belirlenirken en çok kullanılan yöntem aşağıda gösterilmektedir (24).

$$k = (n/2)^{1/2} \quad (12)$$

n: gözlem sayısı

k: küme sayısı

Küme sayısı belirlenirken kullanılan başka bir yöntem ise aşağıda görüldüğü gibidir (25).

$$M = k^2 |W| \quad W = \sum_{j=1}^k \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)(x_{ij} - \bar{x}_j) \quad (13)$$

W: Grup içi kareler toplamı matrisi

k: küme sayısı

2.3. Hiyerarşik Kümeleme Analizi

Hiyerarşik kümeleme teknikleri uygulanırken kaç küme oluşacağı önceden bilinemediği için, kümeleme sürecinin başlangıcında her gözlem bir kümedir ve süreç sonunda ise tüm gözlemler bir kümede toplanır. Süreç 4 aşamalı bir algoritma ile ifade edilebilir (26):

1. n tane gözlem, n tane küme olmak üzere işleme başlanır.
2. En yakın (d_{ij} değeri en küçük olan) iki küme birleştirilir.
3. Küme sayısı bir indirgenerek yinelenmiş uzaklıklar matrisi bulunur.
4. 2 ve 3 nolu adımlar $n-1$ kez tekrarlanır.

şeklinde özetlenebilir. Bu süreçte birden çok gözlemlili kümenin vektör olarak gösterilebilmesi amacıyla değişkenlerin ortalama değerlerinden yeni vektör oluşturmakta ya da bu kümedeki tüm gözlemler ile başka kümedeki gözlemlerin uzaklık ortalamaları da kullanılabilir (26). Hiyerarşik kümeleme, yaygın olarak kullanılan bir veri analiz aracıdır. Amaç, benzer noktadaki grupları birbiriyle birleştiren ikili bir ağaç (dendogram) oluşturmaktır (27).

Hiyerarşik kümeleme sıralı bir süreçtir. Hiyerarşik kümelene için kullanılan yöntemler aglomeratif (kümelenmiş) ve bölücü yaklaşımlar olarak sınıflandırılabilir (1).

Aglomeratif hiyerarşik kümelene, ilk adımda her bir gözlemi kendi bireysel kümesine ayırır, böylece kümelerin ilk sayısı toplam gözlem sayısına eşit olur (28,29). Her kümedeki bireyler n kümeleri ile başlar ve sonra birbirine yakın olan gruplar yeni bir kümelene oluşturmak için birleştirilir. Kümelerin sayısı azalır ve her kümedeki gözlem sayısı artar veya küme büyür. Bu yaklaşım, tüm gözlemleri içeren bir küme ile biter. Her adımda, uzaklık tüm çiftler arasında hesaplanır (1).

Bölücü yaklaşımlar ise, aglomeratif yaklaşımın tersine bir yönde çalışır. Bu yaklaşımda, tüm gözlemleri içeren tek bir kümelene ile başlanır. Algoritma daha sonra kümeyi iki kümeye ayırır. Her adımda iki kümeden biri (gruplar) iki alt gruba ayrılır. Süreç stabilite sağlanıncaya kadar devam eder (1).

Hiyerarşik kümeleme teknikleri uygulanırken kaç küme oluşacağı önceden bilinemez. Öyle ki, kümeleme sürecinin başlangıcında her birey bir kümedir, süreç sonunda ise tüm bireyler bir kümede toplanır. Bu süreçte birden çok gözlemlili kümenin vektör olarak gösterilebilmesi amacıyla değişkenlerin ortalama değerlerinden yeni

vektör oluşturmakta ya da bu kümedeki tüm gözlemler ile başka kümedeki gözlemlerin uzaklık ortalamaları da kullanılabilir (30). En yaygın kullanılan hiyerarşik kümeleme yöntemleri sırasıyla; "Tek Bağlantı", "Ortalama Bağlantı", "Wards Yöntemi" ve "Küme Merkezleri" yöntemleridir (31).

2.3.1. Tek Bağlantı Yöntemi (Nearest Neighbor Method)

En yakın komşu veya minimum yöntem olarak adlandırılır. Bu ölçü, ilk kümeden bir gözlem ile ikinci kümeden bir gözlem arasında bulunan minimum mesafe olarak iki küme arasındaki mesafeyi tanımlar (32,33).

$$d_{mj} = \min\{d_{kj}, d_{lj}\} \quad (14)$$

d_{mj} : m 'inci kümenin j' inci küme ile olan uzaklığı

d_{kj} : k 'inci kümenin j' inci küme ile olan uzaklığı

d_{lj} : l 'inci kümenin j' inci küme ile olan uzaklığı

Tek bağlantı yöntemini kullanmanın bir dezavantajı, bazen kümeler arasında zincirleme oluşturabilmesidir. Bu, birkaç kümenin basitçe bir araya getirilebileceği anlamına gelir. Bu problem çiftler arasındaki en küçük mesafenin dikkate alınan tek değer olması gerçeğinden dolayı tek bağlantı yöntemine özgüdür. Aglomeratif hiyerarşik kümelendirmelerdeki adımlar geri döndürülemez olduğundan, bu zincirleme etkisi küme çözümü üzerinde kötü etkiler yaratabilir (29). Tek bağlantı yöntemi, aykırı değerlere hassastır. Fakat kümelerin yoğunluğundaki farklılıklara karşı dayanıklıdır (34).

2.3.2. Tam Bağlantı Yöntemi (Furthest Neighbor Method)

En uzak komşu veya maksimum yöntem olarak adlandırılır. Bu yöntem, tek bağlantı yöntemine benzerdir, ancak durum çiftleri arasındaki minimum mesafeyi aramak yerine, durum çiftleri arasındaki en uzak mesafeyi dikkate alır (35).

m. kümenin j. küme ile olan uzaklığı:

$$d_{mj} = \max(d_{kj}, d_{lj}) \quad (15)$$

Tam bağlantı yöntemi, büyük kümelendirmeleri yıkabilir, ancak bu durum aykırı değerlerden büyük ölçüde etkilenir (35).

2.3.3. Ortalama Bağlantı Yöntemi (Average Linkage Method)

Tek ve tam bağlantı yöntemlerinin sınırlılıklarının üstesinden gelmek için Sokal ve Michener (1958), gözlem çiftlerinin uzaklık değerinin ortalamasını almayı önermişlerdir (29). Bu yöntemin, kümeler arasındaki mesafenin daha doğru bir değerlendirilmesini sağlamak için bağlantı ölçümleri arasında doğal bir uzlaşmayı temsil etmesi beklenmektedir. Ortalama bağlantı için, ilk kümedeki her gözlem arasındaki ve ikinci kümedeki her gözlem arasındaki mesafeler hesaplanır ve daha sonra ortalaması alınır (32,33).

$$d_{(pq)t} = \frac{\sum_i \sum_k d_{ik}}{N_{(pq)} N_t} \quad (16)$$

$N_{(pq)}$: (pq) kümesindeki eleman sayısı

N_t : t kümesindeki eleman sayısı

Her bağlantı ölçüsü, iki küme arasındaki mesafeyi benzersiz bir şekilde tanımlar. Seçilen bağlantı ölçüsü, kümeleme yöntemi ve kümelennmelerin nasıl bir araya getirildiği üzerinde doğrudan bir etkiye sahip olacaktır (36). Ortalama bağlantı, kümelerin şekline ve boyutuna duyarlı olduğu için tam bağlantı ve tek bağlantı arasındaki doğal bir uyuşmayı temsil eder (34).

2.3.4. Medyan Bağlantı Kümeleme Yöntemi (Median Linkage)

İki küme arasındaki uzaklık hesaplanacağı zaman, küme merkezleri arasındaki uzaklık ağırlıklı olarak hesaplanır(37).

$$d_{(k,l)j} = \frac{(d_{(k,j)} + d_{(l,j)})}{2} - \frac{d_{(k,l)}}{4} \quad (17)$$

$d_{(k,l)j}$: k ve l. kümenin j. küme ile olan uzaklığı

$d_{(k,j)}$: k. kümenin j. kümeyle olan uzaklığı

$d_{(l,j)}$: l. kümenin j. kümeyle olan uzaklığı

$d_{(k,l)}$: k. kümenin l. kümeyle olan uzaklığı

2.3.5. Küme Merkezleri Yöntemi (Centroid Clustering)

İki küme arasındaki uzaklık kümelerin kendi merkezleri arasındaki uzaklık olarak alınır (21). Kümeye yeni birimler eklendiği zaman kümenin merkezi değişir. Bununla birlikte, ortalamaya göre küme merkezi belirlendiği için uç değerlerden etkilenir (38). Kümeler, küme merkezleri arasındaki uzaklığa göre birleştirildiği için küme merkezleri yöntemi denilmektedir (39).

Küme merkezleri yöntemi;

$$d_{(k,l)j} = \frac{(N_k d_{(k,j)} + N_l d_{(l,j)})}{N_k + N_l} - \frac{N_k N_l d_{(k,l)}}{N_k^2 + N_l^2} \quad (18)$$

N_k : k. kümedeki toplam gözlem sayısı

N_l : l. kümedeki toplam gözlem sayısı

$d_{(k,l)j}$: k ve l. kümenin j. küme ile olan uzaklığı

$d_{(k,j)}$: k. kümenin j. kümeye olan uzaklığı

$d_{(l,j)}$: l. kümenin j. kümeye olan uzaklığı

$d_{(k,l)}$: k. kümenin l. kümeye olan uzaklığı

2.4. Hiyerarşik Olmayan Kümeleme Analizi

Küme sayısı konusunda ön bilgi var ise ya da araştırmacı anlamlı olacak küme sayısına karar vermiş ise bu durumda çok uzun zaman alan hiyerarşik yöntemler yerine hiyerarşik olmayan kümeleme yöntemleri tercih edilmektedir. Ayrıca, bu yöntemlerin kuramsal dayanaklarının daha güçlü olması diğer bir tercih sebebidir. Hiyerarşik olmayan kümeleme başlığı altında pek çok yöntemden söz etmek mümkündür. Ancak bunlardan en çok kullanılan iki tanesi "k-ortalama tekniği" ve "en çok olabilirlik tekniği" dir (40).

2.4.1. En Çok Olabilirlik Tekniği

Her bir gözlem en büyük olabilirlik değeri verecek biçimde daha önceden belirlenen kümelere atanmaktadır. Bu yöntem kuramsal dayanağı güçlü bir yöntemdir, ancak yaygın olarak kullanılmamaktadır (40,41).

3. MATERYAL VE METOT

Bu çalışmada, hiyerarşik kümeleme analizi yöntemlerinden olan Ward's yöntemi, hiyerarşik olmayan kümeleme yöntemlerinden k-ortalama yöntemi ve iki aşamalı kümeleme yöntemi kullanılmıştır. Yapılan uygulamada, ülkeler sağlık göstergesi olarak kabul edilen değişkenlere göre kümelenecek olup veri kaynağı olarak Dünya Sağlık Örgütü ve Dünya Bankası web sitelerinde yayınlanmış olan 2013 yılı verileri kullanılmıştır.

3.1. Ward's Yöntemi

Joe H. Ward tarafından 1963 yılında ortaya atılan bu yöntemde, küme içindeki homojenliğin maksimum olması istenir (22). Bir kümenin ortasına düşen gözlemin, aynı kümenin içinde bulunan gözlemlerden ortalama uzaklığını esas alır (21).

Küme merkezinin (ortalamasının) hesaplanması(42);

$$m_i = \frac{1}{n_i} \sum_{x \in G_i} x$$

n_i : G_i kümesindeki gözlem sayısı

Hata kareler toplamı;

$$E = \sum_{k=1}^K \sum_{x \in G_k} \|x_i - m_k\|^2$$

İki kümenin birleşimiyle oluşan küme ile diğer kümeler arasındaki hata kareleri farkı;

$$\Delta E_{ij} = \frac{n_i n_j}{n_i + n_j} \|m_i - m_j\|^2$$

Ward's yöntemi, Centroid ve medyan bağlantı yönteminin ağırlıklı ve karma biçimidir (40).

3.2. K-Ortalama Yöntemi

Tıp uygulamaları için genellikle k-ortalama kümeleme analizi kullanılmaktadır. Veri kümelenebilmesinin, tıbbi veri kümelerindeki yapıyı keşfetmek için etkili bir yöntem

olduğu kanıtlanmıştır (15). Kümeleme algoritmalarının çoğunluğu, her bir gözlemin sadece bir kümeye ait olabileceği anlamına gelen özel kümeler üretir (15,43,44). En basit ve en etkili örtüşen kümeleme yöntemlerinden biri, geleneksel k-ortalama algoritmasının bir uzantısı olan örtüşen k-ortalama (OKM) olarak bilinir (15).

k- Ortalama yöntemi verileri tanımlamak için kümelerin merkezleri olan k prototiplerini kullanır. Hata kareler toplamını minimize ederek belirlenirler (45).

$$J_k = \sum_{k=1}^K \sum_{i \in C_k} (x_i - m_k)^2$$

$x = (x_1, \dots, x_n) \rightarrow$ Veri matrisi

$m_k = \sum_{i \in C_k} x_i / n_k \rightarrow$ C_k kümesinin merkezi

n_k : C_k 'daki noktaların sayısı

Bazı notasyonlar temel bileşenler analizindedir. X , orijinal veri matrisini temsil etmektedir.

$y = (y_1, \dots, y_n)$, $y_i = x_i - \bar{x}$, $\bar{x} = \sum x_i / n$ (merkezi veri matrisi)

$$\sum_i (x_i - \bar{x})(x_i - \bar{x})^T = YY^T \quad (\text{Kovaryans Matrisi})$$

Mac Queen tarafından geliştirilen bu yöntemde, önce araştırmacının ön bilgisine ve tecrübesine dayanarak küme sayısı belirlenir. Sonra her kümenin tipik bir gözlemi seçilir, benzer gözlemler tipik gözlemin etrafında birer birer kümelendirilir. Burada bazı istatistiksel testler kullanılarak her kümeyi oluşturan gözlemlerin değişkenlere göre ortalamalarına bakılır. Güvenilir olması en belirgin üstünlüğüdür. Buna karşılık yorumlaması zordur (21).

3.3. İki Aşamalı Kümeleme Analizi

Hiyerarşik ve k-ortalama kümelenebilirliği n çok büyük olduğunda verimli bir şekilde ölçülemediği için büyük veri tabanlarında iki aşamalı kümeleme analizi kullanılmaktadır (46). Bu analiz, hem kategorik hem de sürekli değişkenler için

kullanılabilir. Ayrıca, uygulamasında üç veya daha fazla kategoriye sahip kategorik değişkenler olduğu zamanda kullanılmaktadır. İki aşamalı kümeleme analizi, hiyerarşik yöntemleri kullanarak alt kümeler oluşturur. Analiz, büyük veri tabanı içeriyorsa, kategorik değişkenlerin dâhil olduğu durumlarda olduğu gibi, iki aşamalı kümeleme analizi önerilir

Süreç iki aşamadan oluşmaktadır (47):

- **İlk Adım:** Gözlemlerin küçük alt kümelere ilk kümeleneşi gerçekleştirilir ve ayrıca bu alt kümeler ayrı gözlemler olarak kabul edilir. Gözlemin önceden oluşturulmuş kümeye ya da yeni bir kümeye katılıp katılmadığına dair karar, uzaklık kriterleri baz alınarak verilir. Bu yeni gözlemlerin gruplandırılması hiyerarşik kümeleme yöntemi ile yapılır. İki aşamalı kümeleme analizinin algoritmasının küme sayısını belirlemesi mümkündür veya küme sayısı önceden atanabilir.
- **İkinci Adım:** Alt kümelerin analiz için temel olduğu gruplamadır ve gereken sayıda kümeye ayrılır. Alt kümelerin sayısı gözlem sayısından önemli ölçüde daha küçük olduğu için, geleneksel gruplama yöntemlerinin kullanımı kolaydır. Daha fazla alt küme varsa, yöntem daha kesindir.

Bu analizde, bir veya daha fazla değişken kategorik ise, log-olabilirlik uzaklık ölçüsü Meila ve Heckerman tarafından geliştirilen bir yöntem kullanılarak gözlemlerin, bu ölçünün en yüksek değerlerine sahip olan kümede gruplandırılması şeklinde kullanılır (48). Tüm değişkenlerin sürekli olması durumunda Öklid uzaklık ölçüsü kullanılır; böylece gözlemler en küçük Öklid mesafesine sahip küme içinde gruplanır. SPSS algoritması, log-olabilirlik yöntemi kategorik ve sürekli değişkenlerle uyumlu olduğu için log-olabilirlik uzaklık ölçüsünde bir azalma kullanır.

Log-olabilirlik uzaklık ölçüsü kullanan iki aşamalı kümeleme analizi yöntemi, sürekli değişkenler için normal dağılım ve kategorik değişkenler için multinom dağılım varsayımını gerektirmektedir. İki aşamalı küme analizi, normallik varsayımına uyulmasa bile iyi sonuçlar verir. Bu analizin bir başka varsayımı, örneklemin büyük (>200) olmasıdır (46). Öklid uzaklığı sadece değişkenlerin tümü sürekli olduğunda bir durumda kullanılabilir. İki küme arasındaki mesafe, onların centroidleri arasındaki Öklid uzaklığı ile tanımlanmaktadır (49).

R ve S kümeleri arasındaki uzaklık (49);

$$d_{(R)(S)} = \xi_R + \xi_S - \xi_{(R,S)}$$
$$\xi_v = -N_v \cdot \left(\left(\sum_{k=1}^{K^A} \frac{1}{2} \cdot \log(\hat{\sigma}_k^2 + \hat{\sigma}_{v,k}^2) \right) + \left(\sum_{k=1}^{K^B} \hat{E}_{v,k} \right) \right)$$
$$\hat{E}_{v,k} = - \sum_{l=1}^{L_k} \left(\frac{N_{v,k,l}}{N_v} \cdot \log \left(\frac{N_{v,k,l}}{N_v} \right) \right)$$

K^A : Analizdeki sürekli değişkenlerin toplam sayısı

K^B : Analizdeki kategorik değişkenlerin toplam sayısı

R^K : k sürekli değişkenlerinin aralığı ya da sırası

N : Veri tabanındaki gözlemlerin sayısı

N_k : k kümedeki gözlemlerin sayısı

$\hat{\sigma}_k^2$: tüm veride k sürekli değişkenlerinin varyans kestirimi

$\hat{\sigma}_{Rk}^2$: R kümesindeki k sürekli değişkenlerin varyans kestirimi

N_{Rkl} : l kategorisini alan k kategorik değişkenlerin R kümesindeki gözlemlerin sayısı

$d_{(R)(S)}$: R ve S kümeleri arasındaki uzaklık

(R, S) : R ve S kümelerinin birleştirilmesiyle oluşan kümeyi temsil eden endeks

Eğer denklemde $\hat{\sigma}_k^2$ dikkate alınmazsa, R ve S kümeleri arasındaki uzaklık iki küme birleştirildiğinde log-olabilirlik uzaklık ölçüsünün azaltılmış değerine eşit olacaktır. $\hat{\sigma}_k^2$ ifadesi, doğal logaritma için tanımsız değerlere ulaşır, $\hat{\sigma}_{vk}^2=0$ olursa, artan sorunun bir çözümü olarak verilir (46).

Bu yöntemin avantajı, Ward's yönteminin, k-ortalama yönteminin gerektirdiği küme sayısı hesaplaması ve karma ölçekli veri setleri için kullanılabilmesidir.

Kümeleme süreci başladığında, kaç kümenin oluşturulması gerektiği veri tabanına bağlıdır. Hiyerarşik kümeleme analizinin özelliği, bir, iki, üç veya daha fazla küme olmak üzere bir dizi çözüm oluşturmaktır. K-ortalama küme algoritması bir dizi

çözüm üretebilmek için birkaç kez (her seferinde farklı kümeler için) gerçekleştirilmelidir.

Küme sayısının otomatik olarak belirlenmesi için SPSS, hiyerarşik kümeleme analizi ile uyumlu iki aşamalı yöntemi geliştirmiştir. İlk adımda, BIC (Schwarz's Bayesian Information Criterion) Bayes bilgi kriteri, ya da AIC (Akaike's Information Criterion) istatistikleri, her farklı küme çözümü için farklı kümelenme sayısı için hesaplanır.

R kümeleri için BIC ve AIC istatistikleri (49);

$$BIC_R = -2 \sum_{i=1}^R \xi_R + m_R \cdot \log(N)$$

$$AIC_R = -2 \sum_{i=1}^R \xi_R + 2 \cdot m_R$$

$$m_R = R \cdot \left\{ 2 \cdot K^A + \sum_{k=1}^K (L_k - 1) \right\}$$

L_k : k kategorik değişkeninde grupların sayısı

Aykırı değerlerin dönüşümü bulunmuyorsa, gözlemleri en yakın kümeye atayarak gözlemlerin kümeye ataması yapılır. Aykırı bir değer dönüşümü varsa, log-olabilirlik mesafesi kullanılır.

Aşırı gözlemlerin normal bir dağılım izlediğini varsayalım. Biri gözlem standartlaştırılmamış kümeye, diğeri ise standartlaştırılmamış kümenin en yakın kümesine atandığında iki ayrı olasılık işlevi hesaplanır. Gözlem daha sonra log-olabilirlik fonksiyonunun en yüksek değerine sahip olan kümeye atanır. Yöntem, kümeden uzaklığı $C = \log(V)$ olan kritik değerden daha küçükse, gözlemin en yakın kümeye (standartlaştırılmamış) atanmasına eşittir.

Burada $V = (\prod_k R_k) \cdot (\prod_m L_m)$ 'dir. Durumların geri kalanında ise gözlem, aykırı değer olarak sınıflandırılır (49).

Öklid uzaklığı kritik değer $C = 2 \cdot \sqrt{\frac{\sum_{l=1}^{K^A} \hat{\sigma}_{k,l}^2}{K^A}}$ den daha küçükse, nesne en yakın kümeye (standartlaştırılmamış) atanır. Aksi halde, aykırı değer olarak sınıflandırılır. Kayıp değerlere izin verilmez. Kayıp değerlerin olduğu gözlemler analizden çıkarılır.

3.4. Veri Analizi

Bu çalışmada kullanılan verilerde sağlık göstergesi olarak;

x₁: toplam doğurganlık hızı (bin kişi başına),

x₂: kaba ölüm hızı (bin kişi başına),

x₃: kaba doğum hızı,

x₄: doğumda beklenen yaşam süresi,

x₅: bebek ölüm hızı,

x₆: beş yaş altı ölüm hızı,

x₇: doktor sayısı,

x₈: hemşire ve ebe sayısı,

x₉: yatak sayısı

değişkenleri ele alınmıştır. Araştırmada kullanılacak olan sağlık göstergeleri değişkenlerine ait veriler DSÖ (Dünya Sağlık Örgütü) ve DB (Dünya Bankası)'nın web sitelerinden alınmış olup, 2017-2016-2015 ve 2014 yıllarına ait verilerde eksiklikler olduğu için 2013 yılına ait tam verileri içermektedir.

Sağlık göstergesi değişkenlerine göre 53 ülkenin tam verisine ulaşılmıştır. Tam verisine ulaşılan ülkelerden 32'si gelişmekte olan, 21'i ise gelişmiş ülkelerdir. Gelişmekte olan ülkeler sırasıyla Arabistan, Arjantin, Bahreyn, Brezilya, Kolombiya, Kosta Rika, Ürdün, Kazakistan, Kırgızistan, Kuveyt, Nikaragua, Tacikistan, Türkmenistan, Türkiye, Özbekistan, Meksika, Jamaika, Ermenistan, Avusturya, Bulgaristan, Bosna Hersek, Gürcistan, Hırvatistan, Macaristan, Moldova, Polonya, Romanya, Rusya, Ukrayna, Şili, Dominik ve Belarus'tur. Gelişmiş ülkeler ise Amerika, Almanya, Azerbaycan, Belçika, Çekya, İsrail, Lüksemburg, İsviçre, Danimarka, İspanya, Estonya, Finlandiya, İzlanda, Litvanya, Letonya, Norveç, Yeni Zellanda, Singapur, Slovakya, Slovenya ve İsveç'tir. Ünelere göre sağlık göstergeleri Tablo-3.1'de gösterilmiştir. Veri setine IBM SPSS Statistics sürüm 25.0 paket programı ile sırasıyla hiyerarşik kümeleme yöntemi olan Ward's yöntemi, hiyerarşik olmayan kümeleme yöntemi olan k-ortalama yöntemi ve iki aşamalı kümeleme yöntemi uygulanmıştır (50).

Çalışmada kullanılan sağlık göstergelerine ilişkin veri seti Tablo 3.1' de sunulmuştur.

Tablo 3.1. Ülkelere Göre Sağlık Göstergeleri Veri Seti

Ülkeler	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉
Arabistan	3.443	5.686	27.191	70.598	29.939	40.006	1.063	2.012	1.145
Arjantin	2.335	7.590	17.717	76.089	11.300	12.700	3.907	4.212	4.900
Ermenistan	1.648	9.727	14.446	74.044	13.800	15.500	2.811	5.465	4.100
Avusturya	1.440	9.400	9.400	81.137	3.200	3.900	4.984	8.022	7.600
Azerbaycan	1.980	5.800	18.300	71.656	26.800	30.400	3.415	7.131	4.700
Belçika	1.750	9.800	11.200	80.588	3.400	4.200	2.959	10.744	6.300
Bulgaristan	1.480	14.400	9.200	74.861	7.800	9.300	3.984	5.378	6.800
Bahreyn	2.098	2.332	15.729	76.482	6.700	7.800	0.929	2.440	2.100
BosnaHersek	1.322	10.533	9.096	76.390	5.500	6.400	1.886	5.767	3.500
Brezilya	1.765	5.971	14.852	74.777	14.900	16.700	1.852	7.444	2.300
İsviçre	1.520	8.000	10.200	82.798	3.800	4.400	4.025	17.607	4.700
Kolombiya	1.921	5.818	16.074	73.847	14.400	16.800	1.766	1.026	1.500
Kosta Rika	1.841	4.767	15.022	79.254	8.400	9.700	1.150	0.796	1.100
Şili	1.702	7.660	11.080	79.282	4.400	5.700	7.366	7.777	5.100
Dominik	1.373	6.775	11.237	79.995	2.500	3.100	2.412	4.085	3.400
Çekya	1.460	10.400	10.200	78.176	2.500	3.100	3.677	8.385	6.500
Almanya	1.420	11.100	8.500	80.490	3.300	4.000	4.039	13.306	8.300
Danimarka	1.670	9.300	10.000	80.300	3.500	4.100	3.648	16.810	3.100
İspanya	1.270	8.300	9.100	83.078	2.900	3.500	3.824	5.161	3.000
Estonya	1.520	11.600	10.300	77.141	2.900	3.600	3.277	5.870	5.000
Finlandiya	1.750	9.500	10.700	80.976	2.200	2.600	3.139	14.809	4.900
Belarus	1.830	9.000	12.100	81.005	3.900	4.600	2.778	8.696	2.800
Gürcistan	1.996	13.131	14.223	72.832	12.200	13.600	4.477	3.898	2.600
Hırvatistan	1.460	11.800	9.400	77.127	4.300	5.000	3.021	6.937	5.900
Macaristan	1.350	12.800	9.000	75.566	4.700	5.600	3.199	6.576	7.000
İzlanda	1.930	6.700	13.400	82.061	1.800	2.400	3.590	16.187	3.200

İsrail	3.030	5.200	21.300	82.056	3.300	4.100	3.539	5.014	3.100
Jamaika	2.072	6.856	17.162	75.523	14.800	17.200	0.482	1.479	1.700
Ürdün	3.560	3.840	28.045	73.883	16.500	19.200	2.591	4.055	1.800
Kazakistan	2.640	7.980	22.730	70.450	13.100	14.700	3.501	8.542	6.700
Kırgızistan	3.100	6.100	27.200	70.202	22.000	24.800	1.886	6.361	4.500
Kuveyt	2.046	2.597	17.894	74.338	8.200	9.500	2.486	6.089	2.200
Amerika	2.124	5.678	17.791	74.640	16.879	20.005	1.915	4.288	1.887
Litvanya	1.590	14.000	10.100	73.915	4.000	5.000	4.268	7.837	7.300
Lüksemburg	1.550	7.000	11.300	81.800	2.300	2.800	2.799	12.267	5.100
Letonya	1.520	14.300	10.200	73.983	5.200	6.100	3.193	5.083	5.800
Moldovya	1.263	11.436	10.849	70.961	14.200	16.500	2.561	5.671	5.800
Meksika	2.270	4.766	19.145	76.579	13.700	15.900	2.071	2.509	1.600
Nikaragua	2.300	4.764	20.788	74.608	17.000	19.900	0.916	1.386	0.900
Norveç	1.780	8.100	11.600	81.751	2.300	2.900	4.303	17.198	3.900
YeniZellanda	2.010	6.650	13.200	81.407	4.900	5.900	2.818	10.644	2.800
Polonya	1.290	10.200	9.700	77.000	4.500	5.200	2.207	5.785	6.500
Romanya	1.460	12.400	9.400	75.063	9.000	10.600	2.669	6.415	6.300
Rusya	1.707	13.000	13.200	70.579	8.100	9.500	4.022	8.794	8.200
Singapur	1.190	4.600	9.300	82.246	2.100	2.700	1.913	5.649	2.100
Slovakya	1.340	9.600	10.100	76.412	5.300	6.400	3.387	6.071	5.800
Slovenya	1.550	9.400	10.200	80.322	2.200	2.700	2.623	8.358	4.600
İsveç	1.890	9.400	11.800	81.956	2.400	2.900	4.119	11.880	2.600
Tacikistan	3.472	5.284	29.998	70.477	33.200	38.200	1.698	4.978	4.800
Türkmenistan	2.958	7.145	26.561	67.375	45.700	54.000	2.266	4.761	7.400
Türkiye	2.108	5.817	17.115	74.977	13.300	15.500	1.755	3.233	2.700
Ukrayna	1.506	14.600	11.100	71.160	8.800	10.200	3.506	7.995	8.800
Özbekistan	2.350	4.800	22.500	70.844	26.000	29.500	2.479	12.561	4.000

4. BULGULAR

Çalışmada kullanılan sağlık göstergelerine ilişkin değişkenlerin tanımlayıcı istatistikleri Tablo 4.1’de gösterilmiştir. Maksimum ve minimum değerler arasında büyük farklılıklar vardır. Ele alınan ülke grubuna ilişkin doğumda beklenen yaşam süresi değişkeninin ortalamasının 76.51 olduğu görülmektedir. 2013 yılı verilerine ait toplam doğurganlık hızı kadın başına maksimum 3.56 iken; her bin doğumdaki bebek ölüm hızı ortalama 9.81’dir.

Tablo 4.1. Sağlık Göstergelerinin Tanımlayıcı İstatistikleri

Değişkenler	N	Minimum	Maksimum	Ortalama	Standart Sapma
X ₁ (toplam doğurganlık hızı)	53	1.190	3.560	1.90471	0.591698
X ₂ (kaba ölüm hızı)	53	2.332	14.600	8.36610	3.164006
X ₃ (kaba doğum hızı)	53	8.500	29.998	14.48951	5.778449
X ₄ (doğumda beklenen yaşam süresi)	53	67.375	83.078	76.51051	4.143289
X ₅ (bebek ölüm hızı)	53	1.800	45.700	9.81167	9.181999
X ₆ (beş yaş altı ölüm hızı)	53	2.400	54.000	11.52097	10.824128
X ₇ (doktor sayısı)	53	0.482	7.366	2.92739	1.195971
X ₈ (hemşire ve ebe sayısı)	53	0.796	17.607	7.19704	4.195043
X ₉ (yatak sayısı)	53	0.900	8.800	4.34776	2.123536

Çalışmada kullanılan veri setine ilk olarak hiyerarşik kümeleme yöntemlerinden biri olan Ward’s yöntemi uygulanmıştır. Ward’s yöntemi az sayıda gözlemin kümelenmesinde kullanılabilir olduğu, optimal ve daha doğru sonuçlar verdiği için tercih edilmiştir. Ward’s yönteminde gözlemler arası uzaklık değerlerinin ölçülmesinde “kare öklit uzaklığı” kullanılmıştır (51). Ancak uzaklık ölçülerinin kullanıldığı aşamalı kümeleme yöntemleri, değişkenler arası birim farklılıklarına çok duyarlı oldukları için verilerin standartlaştırılması uygun olacaktır (51,52). Bu nedenle, ilk olarak verilerin

standartlaştırılmış Z değerleri elde edilmiş ve sonrasında analiz yapılmıştır. Standartlaştırılmış veriler Tablo-4.2’de görülmektedir.

Tablo 4.2. Standardize Edilen Veriler

Ülkeler	Z ₁	Z ₂	Z ₃	Z ₄	Z ₅	Z ₆	Z ₇	Z ₈	Z ₉
Arabistan	2.599	-0.847	2.198	-1.427	2.192	2.632	-1.559	-1.236	-1.508
Arjantin	0.727	-0.245	0.559	-0.102	0.162	0.109	0.819	-0.712	0.260
Ermenistan	-0.434	0.430	-0.008	-0.595	0.434	0.368	-0.097	-0.413	-0,117
Avusturya	-0.785	0.327	-0.881	1.117	-0.720	-0.704	1.720	0.197	1.532
Azerbaycan	0.127	-0.811	0.659	-1.172	1.850	1.744	0.408	-0.016	0.166
Belçika	-0.261	0.453	-0.569	0.984	-0.698	-0.676	0.026	0.846	0.919
Bulgaristan	-0.718	1.907	-0.915	-0.398	-0.219	-0.205	0,883	-0.434	1.155
Bahreyn	0.327	-1.907	0.215	-0.007	-0.339	-0.344	-1.671	-1.134	-1.058
BosnaHersek	-0.985	0.685	-0.933	-0.029	-0.470	-0.473	-0.871	-0.341	-0.399
Brezilya	-0.236	-0.757	0.063	-0.418	0.554	0.478	-0.899	0.059	-0.964
İsviçre	-0.650	-0.116	-0.742	1.517	-0.655	-0.658	0.918	2.481	0.166
Kolombiya	0.028	-0.805	0.274	-0.643	0.500	0.488	-0.971	-1.471	-1.341
Kosta Rika	-0.108	-1.138	0.092	0.662	-0.154	-0.168	-1.486	-1.526	-1.529
Şili	-0.343	-0.223	-0.590	0.669	-0.589	-0.538	3.711	0.138	0.354
Dominik	-0.899	-0.503	-0.563	0.841	-0.796	-0.778	-0.431	-0.742	-0.446
Çekya	-0.752	0.643	-0.742	0.402	-0.796	-0.778	0.627	0.283	1.014
Almanya	-0.819	0.864	-1.037	0.961	-0.709	-0.695	0.929	1.456	1.861
Danimarka	-0.397	0.295	-0.777	0.915	-0.687	-0.686	0.603	2.292	-0.588
İspanya	-1.073	-0.021	-0.933	1.585	-0.753	-0.741	0.750	-0.485	-0.635
Estonya	-0.650	1.022	-0.725	0.152	-0.753	-0.732	0.292	-0.316	0.307
Finlandiya	-0.261	0.358	-0.656	1.078	-0.829	-0.824	0.177	1.815	0.260
Belarus	-0.126	0.200	-0.414	1.085	-0.644	-0.639	-0.125	0.357	-0.729
Gürcistan	0.154	1.506	-0.046	-0.888	0.260	0.192	1.296	-0.786	-0.823
Hırvatistan	-0.752	1.085	-0.881	0.149	-0.600	-0.602	0.078	-0.062	0.731
Macaristan	-0.937	1.401	-0.950	-0.228	-0.557	-0.547	0.227	-0.148	1.249

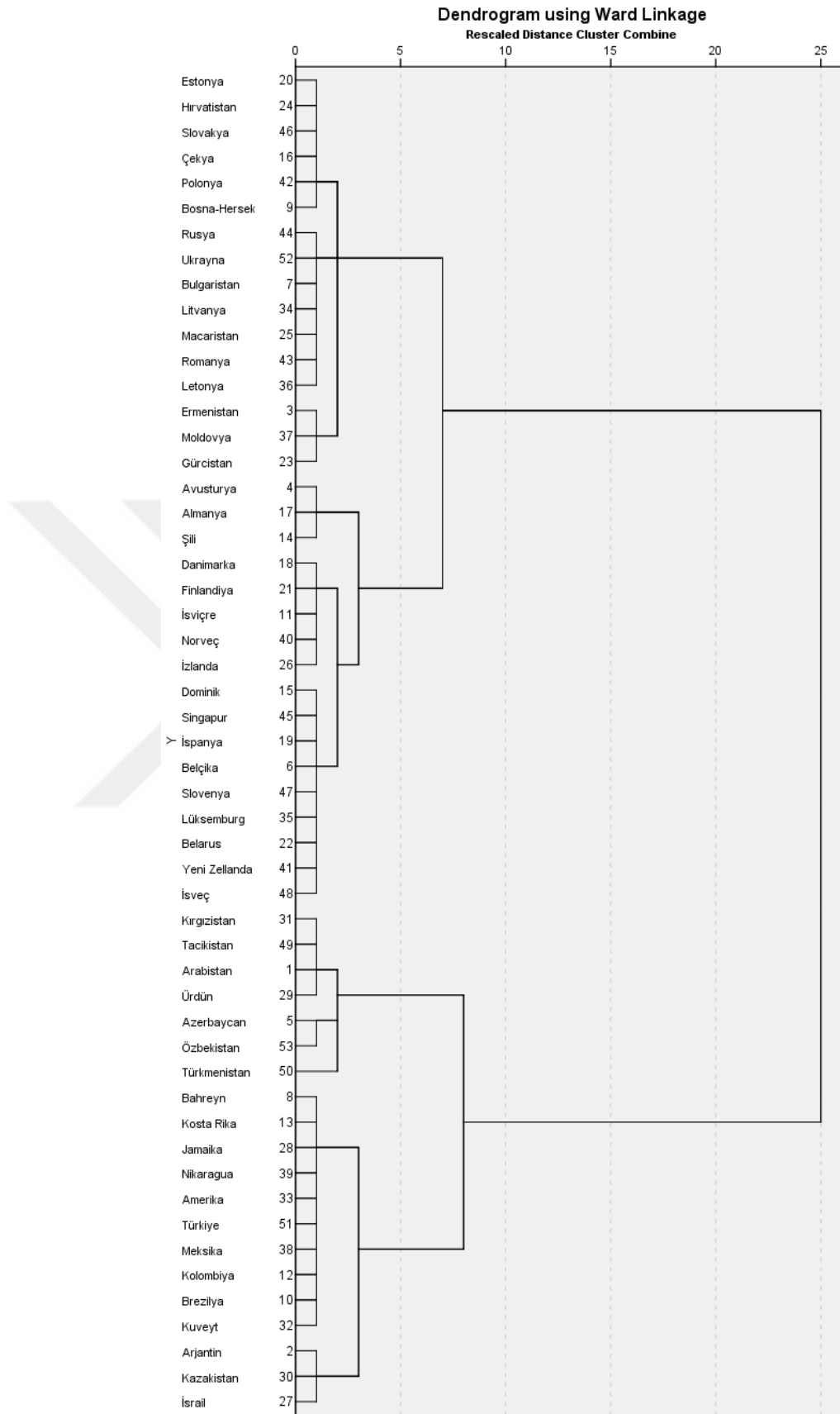
İzlanda	0.043	-0.527	-0.189	1.340	-0.873	-0.843	0.554	2.143	-0.540
İsrail	1.902	-1.001	1.179	1.338	-0.709	-0.686	0.511	-0.520	-0.588
Jamaika	0.283	-0.477	0.462	-0.238	0.543	0.525	-2.045	-1.363	-1.247
Ürdün	2.798	-1.430	2.346	-0.634	0.728	0.709	-0.281	-0.749	-1.200
Kazakistan	1.243	-0.122	1.426	-1.463	0.358	0.294	0.480	0.321	1.108
Kırgızistan	2.020	-0.716	2.200	-1.522	1.327	1.227	-0.871	-0.199	0.072
Kuveyt	0.239	-1.823	0.589	-0.524	-0.176	-0.187	-0.369	-0.264	-1.011
Amerika	0.371	-0.850	0.571	-0.451	0.770	0.784	-0.846	-0.694	-1.159
Litvanya	-0.532	1.781	-0.760	-0.627	-0.633	-0.602	1.121	0.153	1.390
Lüksemburg	-0.599	-0.432	-0.552	1.277	-0.818	-0.806	-0.107	1.209	0.354
Letonya	-0.650	1.875	-0.742	-0.610	-0.502	-0.501	0.222	-0.504	0.684
Moldovya	-1.085	0.970	-0.630	-1.339	0.478	0.460	-0.306	-0.364	0.684
Meksika	0.617	-1.138	0.806	0.017	0.423	0.405	-0.716	-1.118	-1.294
Nikaragua	0.668	-1.138	1.090	-0.459	0.783	0.774	-1.682	-1.385	-1.624
Norveç	-0.211	-0.084	-0.500	1.265	-0.818	-0.796	1.150	2.384	-0.211
YeniZellanda	0.178	-0.542	-0.223	1.182	-0.535	-0.519	-0.091	0.822	-0.729
Polonya	-1.039	0.580	-0.829	0.118	-0.578	-0.584	-0.602	-0.337	1.014
Romanya	-0.752	1.275	-0.881	-0.349	-0.088	-0.085	-0.216	-0.186	0.919
Rusya	-0.334	1.465	-0.223	-1.432	-0.186	-0.187	0.915	0.381	1.814
Singapur	-1.208	-1.190	-0.898	1.384	-0.840	-0.815	-0.848	-0.369	-1.058
Slovakya	-0.954	0.390	-0.760	-0.024	-0.491	-0.473	0.384	-0.268	0.684
Slovenya	-0.599	0.327	-0.742	0.920	-0.829	-0.815	-0.255	0.277	0.119
İsveç	-0.025	0.327	-0.465	1.314	-0.807	-0.796	0.996	1.116	-0.823
Tacikistan	2.649	-0.974	2.684	-1.456	2.547	2.465	-1.028	-0.529	0.213
Türkmenistan	1.780	-0.386	2.089	-2.205	3.909	3.924	-0.553	-0.581	1.437
Türkiye	0.344	-0.806	0.454	-0.370	0.380	0.368	-0.980	-0.945	-0.776
Ukrayna	-0.674	1.970	-0.587	-1.291	-0.110	-0.122	0.484	0.190	2.097
Özbekistan	0.753	-1.127	1.386	-1.368	1.763	1.661	-0.375	1.279	-0.164

Veriler standardize edildikten sonra kare öklit uzaklığı ile Ward's yöntemi uygulanmıştır.

Kare öklit uzaklığına göre ülkelerin sağlık göstergelerine göre ikili karşılaştırmasında en düşük uzaklık değerinin 0.369 ile Estonya ile Hırvatistan arasında olduğu Tablo-4.3'de görülmektedir. Bunu sırası ile, 0.568 uzaklık ölçüsü ile Norveç-İsviçre; 0.576 ile Türkiye-Amerika; 0.736 ile Slovakya-Hırvatistan; 0.809 ile Romanya-Macaristan; 0.812 ile Slovakya-Estonya; 0.829 ile Türkiye-Meksika; 0.832 ile Türkiye-Kolombiya; 0.835 ile Ukrayna -Rusya ve 0.848 ile Hırvatistan-Çekya ülkeleri takip etmektedir. En yüksek uzaklık değerleri incelendiği zaman ise; Türkmenistan ile diğer ülkeler arasında büyük derecede uzaklık olduğu görülmektedir. Örneğin; Türkmenistan'ın Singapur ile 82.752; Norveç ile 81.833; Şili ile 79.986; Danimarka ile 79.208; İzlanda ile 78.930; İsveç ile 77.591; Almanya ile 77.296; Finlandiya ile 75.681; Avusturya ile 75.567; Lüksemburg ile 74.049; Slovenya ile 71.421 ve Dominik ile 71.337'lik bir uzaklık değerine sahip olduğu görülmektedir.

Türkiye ile diğer ülkelerin uzaklık değerleri karşılaştırıldığı zaman ise; Türkiye ile en düşük uzaklık değerine sahip olan ülkenin Amerika olduğu görülmektedir. Bunu; 0,829 uzaklık ölçüsü ile Letonya; 0.832 ile Kolombiya ve 1.586 ile Brezilya takip etmektedir. Buna karşın, Türkiye ile en yüksek uzaklık değerine sahip ülke 38.595 uzaklık ile Türkmenistan olmuştur. Bunu sırasıyla; 29.201 uzaklık ölçüsü ile Şili; 26.815 uzaklık ölçüsü ile Almanya; 24.810 uzaklık ölçüsü ile İsviçre; 23.142 uzaklık ölçüsü ile Norveç ve 22.825 uzaklık ölçüsü ile Avusturya izlemektedir.

Ülkelerin hiyerarşik kümeleme sonuçlarını gösteren Şekil-4.1'deki dendogram incelendiğinde sağlık göstergeleri bakımından birbirine en çok benzerlik gösteren ülkeler 1 birimlik mesafede grup oluştururken, birbirlerine en az benzerlik gösteren ülkeler ise 25 birimlik mesafede bir araya gelmişlerdir. Bazı ülkelerin sağlık göstergeleri birbirlerine çok benzer oldukları için 1 birimlik mesafede grup oluştururken, bazı ülkelerin kendilerine özgü sağlık göstergelerinden dolayı ilk aşamada tek başlarına bir grup gibi görünmektedirler. Estonya-Hırvatistan-Slovakya-Çekya-Polonya-Bosna-Hersek, Rusya-Ukrayna-Bulgaristan-Litvanya-Macaristan-Romanya-Letonya, Ermenistan-Moldovya-Gürcistan, Avusturya-Almanya-Şili, Danimarka-Finlandiya-İsviçre-Norveç-İzlanda, Dominik-Singapur-İspanya-Belçika-Slovenya-Lüksemburg-Belarus-YeniZellanda-İsveç, Kırgızistan-Tacikistan-Arabistan-Ürdün, Azerbaycan-Özbekistan, Bahreyn-KostaRika-Jamaika-Nikaragua-Amerika-Türkiye-Meksika-Kolombiya-Brezilya-Kuveyt, Arjantin-Kazakistan-İsrail arasında çok kuvvetli benzerlikler olduğu için bu ülkeler 1 birimlik mesafede grup oluşturmaktadırlar. Polonya-Moldovya, Finlandiya-Belçika, Arabistan-Azerbaycan-Türkmenistan birbirlerine çok benzeyen ülkeler durumundadır, fakat bu ülkeler dendogramda 2 birimlik mesafede bir araya gelmektedirler. Almanya-İzlanda ülkeleri 3 birimlik mesafede Şili-Danimarka-Finlandiya-İsviçre-Norveç ülkeleri ile bir araya gelmişlerdir. Jamaika-Kazakistan ise 3 birimlik mesafede Nikaragua-Amerika-Türkiye-Meksika-Kolombiya-Brezilya-Kuveyt-Arjantin ülkeleri ile bir araya gelmektedirler. Finlandiya-Ukrayna 7 birimlik mesafede Bulgaristan, Litvanya, Macaristan, Romanya, Letonya, Ermenistan, Moldovya, Gürcistan, Avusturya, Almanya, Şili, Danimarka ülkelerinin bulunduğu gruba dâhil olmuştur. Ürdün-Meksika ise 8 birimlik mesafede Azerbaycan, Özbekistan, Türkmenistan, Bahreyn, Kosta Rika, Jamaika, Nikaragua, Amerika ve Türkiye ülkelerinin bulunduğu gruba dâhil olmuşlardır. Mesafe 25 birim olduğunda ise tüm ülkeler tek bir küme oluşturmaktadır.



Şekil 4.1. Ülkelerin Ağaç Grafığı (Dendrogram)

Şekil-4.1'deki dendograma göre en uygun küme sayısının 2-8 birim aralığında 2 olduğuna karar verilmiştir. Dendograma göre oluşan kümeler Tablo-4.4'de görülmektedir.

Tablo 4.4. Ward's Yöntemi Sonucunda Kümelerde Yer Alan Ülkeler

Küme-1	Küme-2
Estonya, Hırvatistan, Slovakya, Çekya, Polonya, Bosna-Hersek, Rusya, Ukrayna, Bulgaristan, Litvanya, Macaristan, Romanya, Letonya, Ermenistan, Moldovya, Gürcistan, Avusturya, Almanya, Şili, Danimarka, Finlandiya, İsviçre, Norveç, İzlanda, Dominik, Singapur, İspanya, Belçika, Slovenya, Lüksemburg, Belarus, Yeni Zellanda, İsveç	Arabistan, Ürdün, Azerbaycan, Özbekistan, Türkmenistan, Kırgızistan, Tacikistan, Bahreyn, Kosta Rika, Jamaika, Nikaragua, Amerika, Türkiye, Meksika, Kolombiya, Brezilya, Kuveyt, Arjantin, Kazakistan, İsrail

Tablo 4.5. Kümelerin Oluşması Aglomeratif Tablo (Birleştirme Tablosu)

Aşama	Birleştirilmiş Kümeler		Katsayılar	İlk Kümeleşmenin Görüldüğü Aşamalar		Sonraki Aşamalar
	Küme 1	Küme 2		Küme 1	Küme 2	
1	Estonya	Hırvatistan	0.184	0	0	6
2	İsviçre	Norveç	0.468	0	0	16
3	Amerika	Türkiye	0.756	0	0	9
4	Macaristan	Romanya	1.161	0	0	10
5	Rusya	Ukrayna	1.578	0	0	36
6	Estonya	Slovakya	2.033	1	0	11
7	Bulgaristan	Litvanya	2.489	0	0	25
8	Belarus	YeniZellanda	2.955	0	0	23
9	Amerika	Meksika	3.435	3	0	14
10	Macaristan	Letonya	3.999	4	0	25
11	Çekya	Estonya	4.585	0	6	20
12	Danimarka	Finlandiya	5.200	0	0	21
13	Belçika	Slovenya	5.822	0	0	18
14	Kolombiya	Amerika	6.448	0	9	26
15	Jamaika	Nikaragua	7.159	0	0	33
16	İsviçre	İzlanda	7.966	2	0	21
17	Dominik	Singapur	8.799	0	0	28
18	Belçika	Lüksemburg	9.648	13	0	34
19	Bahreyn	KostaRika	10.507	0	0	38
20	Çekya	Polonya	11.462	11	0	31
21	İsviçre	Danimarka	12.474	16	12	47
22	Ermenistan	Moldovya	13.651	0	0	37
23	Belarus	İsveç	14.921	8	0	34
24	Avusturya	Almanya	16.250	0	0	40
25	Bulgaristan	Macaristan	17.600	7	10	36
26	Brezilya	Kolombiya	19.080	0	14	30
27	Azerbaycan	Özbekistan	20.814	0	0	45
28	Dominik	İspanya	22.749	17	0	42
29	Kırgızistan	Tacikistan	24.686	0	0	35
30	Brezilya	Kuveyt	26.786	26	0	33
31	BosnaHersek	Çekya	28.916	0	20	44
32	Arjantin	Kazakistan	31.344	0	0	41
33	Brezilya	Jamaika	34.352	30	15	38
34	Belçika	Belarus	37.364	18	23	42
35	Arabistan	Kırgızistan	40.479	0	29	39
36	Bulgaristan	Rusya	43.766	25	5	43
37	Ermenistan	Gürcistan	47.285	22	0	43
38	Bahreyn	Brezilya	51.248	19	33	49
39	Arabistan	Kazakistan	55.900	35	0	46
40	Avusturya	Şili	62.012	24	0	48
41	Arjantin	İsrail	68.350	32	0	49
42	Belçika	Dominik	74.889	34	28	47
43	Ermenistan	Bulgaristan	83.355	37	36	44
44	Ermenistan	BosnaHersek	91.950	43	31	50
45	Azerbaycan	Türkmenistan	103.523	27	0	46
46	Arabistan	Azerbaycan	116.322	39	45	51
47	Belçika	İsviçre	129.591	42	21	48
48	Avusturya	Belçika	145.506	40	47	50
49	Arjantin	Bahreyn	163.730	41	38	51
50	Ermenistan	Avusturya	216.351	44	48	52
51	Arabistan	Arjantin	272.930	46	49	52
52	Arabistan	Ermenistan	468.000	51	50	0

Aglomeratif tabloda sađlık gstergeleri deđiřkenlerine bađlı katsayıya gre birbirlerine en ok benzeyen lkeler eřleřtirilir. Agglomeratif tabloda n-1 adet ařama vardır. Buna gre lkeler ve 53-1=52 adet ařama sz konusudur. Bu tablo ile hangi lkenin hangi ařamada hangi lke ile kmelendiđi grlebilmektedir. Birinci sırada birbirine en ok benzeyen lkeler bir kme oluřtururken, daha sonra bu kmeye yeni bir lke veya bařka lkelerin oluřturduđu yeni bir kme eklenir. Bu ynteme gre; kme-1'de Estonya ile kme-2' de Hırvatistan birbirine en yakın iki lke olarak grnmektedir. Bu iki lke 6. ařamada aralarına Slovakya'yı almıřlardır. Bir bařka ifadeyle; sađlık gstergeleri bakımından birbirine en yakın olan lkeler Estonya ile Hırvatistan'dır. Daha sonra ise İsvire ve Norve yakın bulunmuřtur. Birbirine en uzak lkeler ise Arabistan-Ermenistan ve Arabistan-Arjantin lkeleridir. Katsayılardan da anlařılacađı zere, sađlık gstergelerine gre birbirine en yakın olan lkeler ilk basamaklarda yer almaktadırlar. Birbirlerine daha az benzeyen lkeler ise daha sonraki ařamalarda yer almaktadırlar.

lkelerin sađlık gstergelerine gre iki kme oluřmasına karar verildikten sonra kme sayısının bilindiđi kmeleme yntemi olan k-ortalama kmeleme analizi uygulanmıřtır. Tablo-4.6'de k-ortalama kmeleme yntemi sonucunda hangi lkenin hangi kmede olduđu ve uzaklık deđerleri gsterilmektedir.

Tablo 4.6. K-Ortalama Yöntemi Sonucunda Elde Edilen Kümeler ve Uzaklıkları

Ülkeler	Küme	Uzaklık	Ülkeler	Küme	Uzaklık
Arabistan	1	3.377	Jamaika	1	2.027
Arjantin	1	2.280	Ürdün	1	2.421
Ermenistan	2	2.064	Kazakistan	1	2.725
Avusturya	2	1.955	Kırgızistan	1	2.073
Azerbaycan	1	2.267	Kuveyt	1	2.041
Belçika	2	1.101	Amerika	1	0.982
Bulgaristan	2	2.049	Litvanya	2	2.026
Bahreyn	1	2.599	Lüksemburg	2	1.712
BosnaHersek	2	1.800	Letonya	2	1.911
Brezilya	1	1.843	Moldovya	2	2.558
İsviçre	2	2.544	Meksika	1	1.319
Kolombiya	1	1.755	Nikaragua	1	1.645
Kosta Rika	1	2.850	Norveç	2	2.523
Şili	2	3.407	YeniZellanda	2	2.070
Dominik	2	2.067	Polonya	2	1.522
Çekya	2	0.806	Romanya	2	1.597
Almanya	2	2.079	Rusya	2	2.623
Danimarka	2	2.244	Singapur	2	2.994
İspanya	2	2.024	Slovakya	2	0.936
Estonya	2	0.942	Slovenya	2	1.038
Finlandiya	2	1.711	İsveç	2	1.925
Belarus	2	1.559	Tacikistan	1	3.515
Gürcistan	2	2.842	Türkmenistan	1	5.230
Hırvatistan	2	0.937	Türkiye	1	1.228
Macaristan	2	1.547	Ukrayna	2	2.853
İzlanda	2	2.618	Özbekistan	1	2.483
İsrail	1	3.337			

Küme ilişkileri tablosuna göre oluşan kümeler Tablo-4.7’de görülmektedir.

Tablo 4.7. K-Ortalama Yöntemi Sonucunda Kümelerde Yer Alan Ülkeler

Küme-1	Küme-2
Arabistan, Arjantin, Azerbaycan, Bahreyn, Brezilya, Kolombiya, Kosta Rika, İsrail, Jamaika, Ürdün, Kazakistan, Kırgızistan, Kuveyt, Amerika, Meksika, Nikaragua, Tacikistan, Türkmenistan, Türkiye, Özbekistan	Ermenistan, Avusturya, Belçika, Bulgaristan, Bosna Hersek, İsviçre, Şili, Dominik, Çekya, Almanya, Danimarka, İspanya, Estonya, Finlandiya, Belarus, Gürcistan, Hırvatistan, Macaristan, İzlanda, Litvanya, Lüksemburg, Letonya, Moldova, Norveç, Yeni Zellanda, Polonya, Romanya, Rusya, Singapur, Slovakya, Slovenya, İsveç, Ukrayna

Kümeleme analizinde kullanılan ülkelerin sağlık göstergelerinin oluşan kümeler ile birlikte ortalamaları Tablo 4.8’de verilmiştir.

Tablo 4.8. Son Küme Merkezleri

Sağlık Göstergeleri*	Küme-1	Küme-2
X ₁ (toplam doğurganlık hızı)	0.95645	-0.57967
X ₂ (kaba ölüm hızı)	-0.92495	0.56057
X ₃ (kaba doğum hızı)	1.06409	-0.64672
X ₄ (doğumda beklenen yaşam süresi)	-0.62214	0.37705
X ₅ (bebek ölüm hızı)	0.87064	-0.52766
X ₆ (beş yaş altı ölüm hızı)	0.86008	-0.52126
X ₇ (doktor sayısı)	-0.70571	0.42770
X ₈ (hemşire ve ebe sayısı)	-0.63914	0.38736
X ₉ (yatak sayısı)	-0.61039	0.36994

* Sağlık göstergelerine ait verilerin z değerleri ile gerçekleştirilmiştir.

Tablo-4.8’ de verilen sağlık göstergelerinin iki kümedeki ortalamaları incelendiği zaman birinci kümeyi oluşturan ülkeler x₂ (kaba ölüm hızı), x₇ (doktor sayısı), x₈ (hemşire ve ebe sayısı), x₄ (doğumda beklenen yaşam süresi) ve x₉ (yatak sayısı)

ortalamalarının en düşük, x_3 (kaba doğum hızı), x_1 (toplam doğurganlık hızı), x_5 (bebek ölüm hızı) ve x_6 (beş yaş altı ölüm hızı) ortalamalarının ise en yüksek olduğu ülkelerdir. İkinci kümeyi oluşturan ülkelerde ise x_2 (kaba ölüm hızı), x_7 (doktor sayısı), x_8 (hemşire ve ebe sayısı), x_4 (doğumda beklenen yaşam süresi) ve x_9 (yatak sayısı) en yüksek iken, x_3 (kaba doğum hızı), x_1 (toplam doğurganlık hızı), x_5 (bebek ölüm hızı) ve x_6 (beş yaş altı ölüm hızı) en düşük olduğu görülmektedir.

Son küme merkezleri arasındaki uzaklıklara ait bilgiler ise Tablo 4.9'da görülmektedir.

Tablo 4.9. Son Küme Merkezleri Arasındaki Uzaklıklar

Kümeler	1	2
1		3.958
2	3.958	

Son küme merkezleri arasındaki uzaklığa ait Tablo-4.9 incelendiği zaman birinci ve ikinci küme arasındaki uzaklığın 3.958 olduğu görülmektedir.

Kümeleme analizinde kullanılan ülkelere ait sağlık göstergelerinin kümeler itibari ile farklılığının öğrenilmesi için ANOVA testi yapılmıştır. İki küme için k-ortalama yöntemi uygulanan değişkenlerin ANOVA tablosu anlamlılık değerleri Tablo-4.10'da gösterilmiştir.

Tablo 4.10. ANOVA tablosu

	Küme		Hata		F	p
	Kareler	Sd	Kareler	Sd		
	Ortalaması		Ortalaması			
X ₁ (toplam doğurganlık hızı)	29.384	1	0.443	51	66.264	<0.001
X ₂ (kaba ölüm hızı)	27.481	1	0.481	51	57.159	<0.001
X ₃ (kaba doğum hızı)	36.576	1	0.302	51	120.940	<0.001
X ₄ (doğumda beklenen yaşam süresi)	12.433	1	0.776	51	16.025	<0.001
X ₅ (bebek ölüm hızı)	24.349	1	0.542	51	44.908	<0.001
X ₆ (beş yaş altı ölüm hızı)	23.761	1	0.554	51	42.913	<0.001
X ₇ (doktor sayısı)	15.997	1	0.706	51	22.661	<0.001
X ₈ (hemşire ve ebe sayısı)	13.122	1	0.762	51	17.213	<0.001
X ₉ (yatak sayısı)	11.968	1	0.785	51	15.247	<0.001

* Sağlık göstergelerine ait verilerin z değerleri ile gerçekleştirilmiştir.

ANOVA tablosu sağlık göstergesi değişkenlerinin kümelere ayırmada anlamlı olup olmadıklarının görülmesi amacıyla kullanılmaktadır. Tablo-4.10 incelendiği zaman ülkelerin iki kümede kümelenmesinde seçilen sağlık göstergesi değişkenlerinin kümelere ayırmada anlamlı oldukları ($p < 0.05$) gözlenmektedir. F değerlerine bakıldığı zaman kümelere ayırmada en etkili sağlık göstergesi değişkeninin x_3 ($F=120.940$) olduğu görülmektedir. En az etkili değişkenler x_4 ($F=16.025$) ve x_9 ($F=15.247$) değişkenleridir. Bu sonucun ortaya çıkmasının sebebi ise, kümeler arası farkın ençoklanmasıdır. Yani, kümelerdeki ülkelerin dağılımı tesadüfi değildir.

Son olarak iki aşamalı kümeleme analizi yöntemi uygulanmıştır. İki aşamalı kümeleme analizi sonucunda BIC kriterine göre küme sayısı 2 olarak belirlenmiştir.

Tablo 4.11. İki Aşamalı Kümeleme Yöntemi Sonucu Oluşan BIC Kriter Tablosu

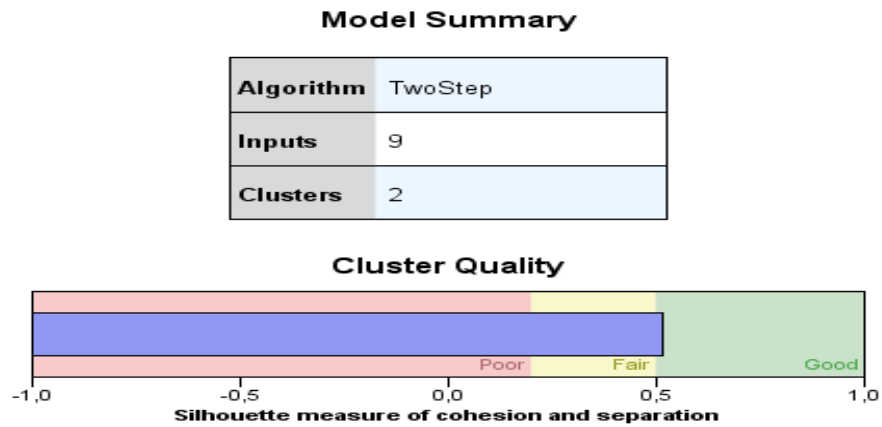
Küme Sayısı	Schwarz'ın Bayesyen Kriteri (BIC)	BIC Değişimi	Oransal BIC Değişimi	Uzaklık Ölçülerinin Oranı
1	397.575			
2	346.500	-51.075	1.000	3.240
3	380.140	33.641	-0.659	1.123
4	417.929	37.789	-0.740	1.942
5	472.051	54.121	-1.060	1.630
6	532.877	60.826	-1.191	1.183
7	595.350	62.473	-1.223	1.035
8	658.130	62.780	-1.229	1.090
9	721.625	63.494	-1.243	1.025
10	785.310	63.686	-1.247	1.161
11	850.074	64.763	-1.268	1.260
12	916.221	66.147	-1.295	1.139
13	983.019	66.798	-1.308	1.082
14	105.171	67.152	-1.315	1.032
15	111.458	67.287	-1.317	1.162

Bu iki küme sırasıyla 33 (%62.3) ve 20 (%37.7) ülkeden oluşmaktadır. Oluşan kümelerin gözlem sayısı bakımından dengeli olduğu söylenebilir.

Tablo 4.12. Küme Merkezleri (Centroids)

Değişkenler		Küme-1	Küme-2	Birleştirilmiş
X ₁ (toplam doğurganlık hızı)	\bar{x}	2.47064	1.56173	1.90471
	σ	0.5794290	0.2192920	0.591698
X ₂ (kaba ölüm hızı)	\bar{x}	5.43955	10.13976	8,36610
	σ	1.439912	2.537642	3.164006
X ₃ (kaba doğum hızı)	\bar{x}	20.65565	10.75245	14.48951
	σ	4.802155	1.549809	5.778449
X ₄ (doğumda beklenen yaşam süresi)	\bar{x}	73.93281	78.07275	76.51051
	σ	3.439337	3.768677	4.143289
X ₅ (bebek ölüm hızı)	\bar{x}	17.80592	4.9667	9.81167
	σ	10.198905	3.330415	9.181999
X ₆ (beş yaş altı ölüm hızı)	\bar{x}	20.83057	5.87879	11.52097
	σ	12.271479	3.738796	10.824128
X ₇ (doktor sayısı)	\bar{x}	2.08338	3.43891	2.92739
	σ	0.959353	1.030920	1.195971
X ₈ (hemşire ve ebe sayısı)	\bar{x}	4.51581	8.82203	7.19704
	σ	2.941478	4.030365	4.195043
X ₉ (yatak sayısı)	\bar{x}	3.05157	5.13333	4.34776
	σ	1.892763	1.874611	2.123536

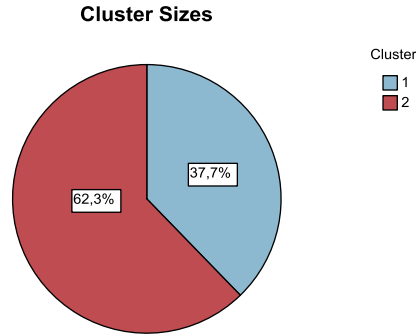
Kümeleme kalitesi ölçütlerinden olan Silhoutte ölçütüne göre kümelemedeki başarı adil (fair) olarak nitelendirilmektedir.

**Şekil 4.2.** Model Özeti ve Kümeleme Kalitesi

Ülkelerin sağlık göstergelerini 2 kümede ele almak kümelemedeki başarıyı artırmıştır.

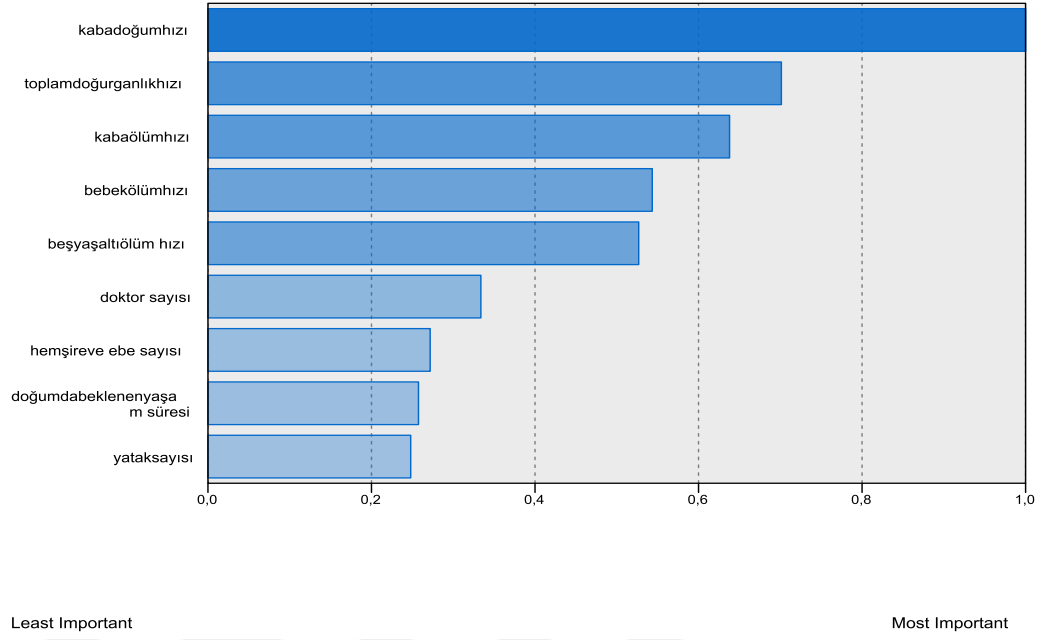
İki aşamalı kümeleme yöntemi sonucunda oluşan kümeler ve sağlık göstergesi değişkenlerinin ortalamaları ve ülkelerin kümelere ayrılmasında kullanılan oranların önem derecesi Şekil-4.3 ve Şekil-4.4' de görülmektedir. Küme-1 ülkelerin % 37.7'sini, Küme-2 ise %62.3'ünü kapsamaktadır. Kümelerin oluşmasında en belirleyici değişken kaba doğum hızı (x_3) olmuştur. Doktor sayısı (x_7), hemşire ve ebe sayısı (x_8), doğumda beklenen yaşam süresi (x_4) ve yatak sayısı (x_9) ise en az etkili olan değişkenlerdir. En düşük kaba ölüm hızının (x_2) 5.44 ile birinci kümede olduğu, en yüksek kaba ölüm hızının ise 10.14 ile ikinci kümede olduğu görülmektedir. Ülkeden ülkeye geçişmekle beraber genel anlamda belirtmek gerekirse, tüm sağlık göstergesi değişkenleri dikkate alınarak kaba doğum hızı (x_3) yüksek, kaba ölüm hızı (x_2) düşük olan birinci kümenin daha iyi bir küme yapısı gösterdiği söylenebilir.

Bu yöntemle oluşan kümeler ve büyüklükleri, küme ortalamaları ve sağlık göstergesi değişkenlerinin önem dereceleri ve sağlık göstergesi değişkenlerinin önem derecelerinin grafikleri sırasıyla Şekil-4.3, Şekil-4.4 ve Şekil-4.5'de gösterilmiştir.



Size of Smallest Cluster	20 (37,7%)
Size of Largest Cluster	33 (62,3%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	1,65

Predictor Importance




Şekil 4.3. Oluşan Küme Büyüklükleri

Clusters

Input (Predictor) Importance

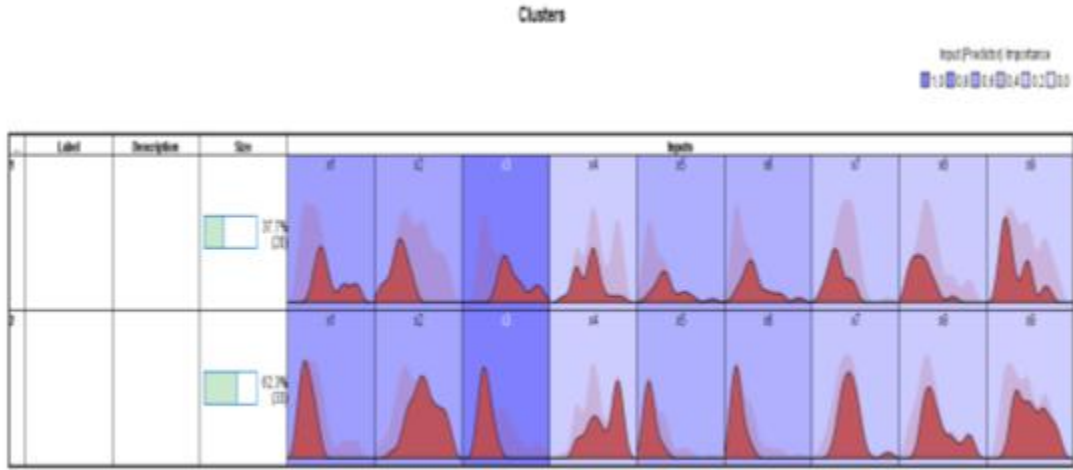
1,0 0,8 0,6 0,4 0,2 0,0

Cluster	2	1
Label		
Description		
Size	 62,3% (33)	 37,7% (20)
Inputs	kabadoğumhızı 10,75	kabadoğumhızı 20,66
	toplamdoğurganlığı zı	toplamdoğurganlığı zı
	kabaölümhızı 10,14	kabaölümhızı 5,44
	bebekölümhızı 4,97	bebekölümhızı 17,81
	beşyaşaltıölüm hızı 5,88	beşyaşaltıölüm hızı 20,83
	doktor sayısı 3,44	doktor sayısı 2,08
	hemşireve ebe sayısı 8,82	hemşireve ebe sayısı 4,52
	doğumdabeklenenya şam süresi	doğumdabeklenenya şam süresi
	yataksayısı 5,13	yataksayısı 3,05

Şekil 4.4. İki Aşamalı Yönteme Göre Oluşan Küme Ortalamaları ve Sağlık Göstergesi Değişkenlerinin Önem Dereceleri

İki aşamalı kümeleme yöntemine göre oluşan kümeler, k-ortalama yöntemine göre oluşan kümeler ile karşılaştırıldığı zaman, aynı oldukları (küme 1: 20, Küme 2: 33)

görülmektedir. Bu durum her iki yöntemin de tercih edilebileceğini göstermektedir. Bu iki yöntemle oluşan kümelerdeki ülkeler ise Ward's yönteminde yer değiştirerek küme oluşturmuşlardır (Küme 1: 33, Küme 2: 20).



Şekil 4.5. Kümelerdeki Sağlık Göstergesi Değişkenlerinin Önem Derecelerinin Grafikleri

Bu yöntem sonucunda oluşan kümeler ve kümelerdeki ülkeler Tablo 4.13'de gösterilmiştir.

Tablo 4.13. İki Aşamalı Kümeleme Analizi Sonucunda Oluşan Kümeler

Küme-1	Küme-2
Arabistan, Arjantin, Azerbaycan, Bahreyn, Brezilya, Kolombiya, Kosta Rika, İsrail, Jamaika, Ürdün, Kazakistan, Kırgızistan, Kuveyt, Amerika, Meksika, Nikaragua, Tacikistan, Türkmenistan, Türkiye, Özbekistan	Ermenistan, Avusturya, Belçika, Bulgaristan, Bosna Hersek, İsviçre, Şili, Dominik, Çekya, Almanya, Danimarka, İspanya, Estonya, Finlandiya, Belarus, Gürcistan, Hırvatistan, Macaristan, İzlanda, Litvanya, Lüksemburg, Letonya, Moldova, Norveç, Yeni Zellanda, Polonya, Romanya, Rusya, Singapur, Slovakya, Slovenya, İsveç, Ukrayna

5. TARTIŞMA

Kümeleme analizi, veri grupları arasındaki uzaklıkları dikkate alarak, küme grupları yaratmaya çalışmaktadır. Ülkelerin sağlık göstergelerinin incelendiği bu çalışmada amaç, sağlık göstergesi açısından hangi ülkelerin birbirine benzer ve hangi ülkelerin birbirinden farklılaştığını kümeleme analizi yöntemlerini kullanarak incelemektir.

Çalışmada 53 ülke belirlenen sağlık göstergelerine göre sınıflandırılmıştır. Sağlık göstergesi olarak; toplam doğurganlık hızı (bin kişi başına), kaba ölüm hızı (bin kişi başına), kaba doğum hızı, doğumda beklenen yaşam süresi, bebek ölüm hızı, beş yaş altı ölüm hızı, doktor sayısı, hemşire ve ebe sayısı, yatak sayısı değişkenleri kullanılmıştır.

Bu çalışmada ülkeler Ward's yöntemi, k-ortalama yöntemi ve iki aşamalı kümeleme yöntemlerine göre kümelendirilmiştir. İlk olarak Ward's yöntemi uygulanmış ve uzaklık ölçüsü olarak öklid uzaklığı kullanılmıştır. Ülkelerin iki kümeye ayrılması uygun görülmüştür. Daha sonra ise k-ortalama yöntemine göre Türkiye, Amerika, Azerbaycan, İsrail, Arabistan, Arjantin, Bahreyn, Brezilya, Kolombiya, Kosta Rika, Ürdün, Kazakistan, Kırgızistan, Kuveyt, Nikaragua, Tacikistan, Türkmenistan, Türkiye, Özbekistan, Meksika ve Jamaika'nın birinci kümeyi, Ermenistan, Avusturya, Bulgaristan, Bosna Hersek, Gürcistan, Hırvatistan, Macaristan, Moldovya, Polonya, Romanya, Rusya, Ukrayna, Şili, Dominik, Belarus, Belçika, İsviçre, Çekya, Lüksemburg, Almanya, Danimarka, İspanya, Estonya, Finlandiya, İzlanda, Litvanya, Letonya, Norveç, Yeni Zelanda, Singapur, Slovakya, Slovenya ve İsveç ise ikinci kümeyi oluşturduğu görülmüştür. Ülkelerin kümeleneşmesi için kullanılan sağlık göstergesi değişkenlerinin tamamının önemli düzeyde etkin oldukları ($p < 0.05$) görülmüştür. Hem Ward's yöntemi, hem k-ortalama yöntemi hem de iki aşamalı yöntemde aynı kümelerin oluştuğu görülmüştür.

Türkiye gelişmekte olan bir ülke olmakla beraber, bu çalışmada dikkate aldığımız sağlık göstergesi değişkenlerine göre Türkiye gelişmiş ülkeler kümesinde yer almıştır.

6. SONUÇ VE ÖNERİLER

Bu çalışmada ülkelerin sağlık göstergelerine göre sınıflandırılmasında kümeleme analizi yöntemlerinin kullanılabilirlikleri test edilmiştir. Bu amaç doğrultusunda çalışmada kullanılan kümeleme analizi yöntemlerinin aşamaları ayrıntılı bir şekilde ortaya konulmuştur. Yapılan çalışma ile aynı zamanda sağlık göstergesi değişkenlerine bağlı olarak çok sayıda ülkenin daha az sayıda ve anlamlı kümelere bölüdüğü görülmüştür.

Çalışmanın sonuçlarına göre Ward's yönteminde iki küme oluşmuştur. Birinci kümede 33 ülke, ikinci kümede ise 20 ülke bir araya gelmiştir. Birinci kümede gelişmekte olan 15 ülke ve gelişmiş 18 ülke, ikinci kümede ise gelişmekte olan 17 ülke ile gelişmiş 3 ülke yer almaktadır.

K-ortalama yönteminde elde edilen sonuçlara göre, birinci kümede gelişmekte olan 17 ülke ile gelişmiş olan 3 ülke olmak üzere toplam 20 ülke bulunmaktadır. İkinci kümede ise gelişmekte olan 15 ülke ile gelişmiş olan 18 ülke olmak üzere toplam 33 ülke bulunmaktadır. Birinci kümede bulunan gelişmekte olan ülkeler sırasıyla Arabistan, Arjantin, Bahreyn, Brezilya, Kolombiya, Kosta Rika, Ürdün, Kazakistan, Kırgızistan, Kuveyt, Nikaragua, Tacikistan, Türkmenistan, Türkiye, Özbekistan, Meksika ve Jamaika'dır. Birinci kümede bulunan gelişmiş ülkeler ise Amerika, Azerbaycan ve İsrail'dir. Bu ülkelerin bir arada toplanması, sağlık göstergeleri bakımından birbirlerine benzer ülkeler olduklarını göstermektedir. Gelişmekte olan ülkelerin gelişmiş ülkelere yaklaştıkları söylenebilir. İkinci kümede bulunan gelişmekte olan ülkeler ise Ermenistan, Avusturya, Bulgaristan, Bosna Hersek, Gürcistan, Hırvatistan, Macaristan, Moldovya, Polonya, Romanya, Rusya, Ukrayna, Şili, Dominik ve Belarus' dur. İkinci kümede bulunan gelişmiş ülkeler Belçika, İsviçre, Çekya, Lüksemburg, Almanya, Danimarka, İspanya, Estonya, Finlandiya, İzlanda, Litvanya, Letonya, Norveç, Yeni Zeland, Singapur, Slovakya, Slovenya ve İsveç' dir. Ayrıca bu yöntemde ANOVA tablosuyla ülkelerin iki kümeye ayrılmasında etkili olan sağlık göstergesi değişkenlerinin istatistiksel olarak anlamlı oldukları görülmüştür.

İki aşamalı kümeleme yönteminde ise, iki küme oluşmuştur. Birinci kümede gelişmekte olan 17 ülke ile gelişmiş olan 3 ülke olmak üzere toplam 20 ülke bulunmaktadır. İkinci kümede ise gelişmekte olan 15 ülke ile gelişmiş olan 18 ülke

olmak üzere toplam 33 ülke bulunmaktadır. Birinci kümede geliřmekte olan ülkelerin geliřmiř ülkelere yaklařtıkları söylenebilir. İkinci kümede ise yine geliřmekte olan ülkelerin geliřmiř ülkelere yaklařtıkları ya da geliřmiř ülkelerin gerileme sürecine geçebilecekleri söylenebilir.

Uygulanan kümeleme analizi yöntemleri sonucunda üç yöntemde aynı sonucu verdiđi gözlenmiştir. Ancak, iki aşamalı kümeleme yöntemi gözlem sayısının fazla olması durumunda, kümelemede dikkate alınan deđişkenler hem nitel hem de nicel olduđunda kullanılabilir olması nedeniyle bazı durumlarda diđer yöntemlere göre avantajlı konuma gelmektedir.

Bu çalışmada 3 yöntem de aynı sonucu vermekle beraber, deđişken sayısı ve gözlem sayısı deđiřtiđinde ya da farklı olduđunda yöntemler farklı sonuçlar verebilir.

KAYNAKLAR

1. Alkarkhi AFM, Alqaraghuli WAA. *Cluster Analysis*. Easy Statistics for Food Science with R. 2019, 177-86.
2. Sokal RR, Sneath PHA. *Principles of Numerical Taxonomy*. San Francisco 1963.
3. Çelik Ş. Classification of Provinces in Turkey according to Health Indicators by Cluster Analysis, *Doğuş Üniversitesi Dergisi* 2013, 14(2): 175-94.
4. Nugent R, Meila M. An overview of clustering applied to molecular biology. *Statistical Methods in Molecular Biology* 2010, 369-404.
5. Wiwie C, Baumbach J, Röttger R. Comparing the performance of biomedical clustering methods. *Nature Methods* 2015, 12(11): 1033-38.
6. Naik D, Shah P. A Review on image segmentation clustering algorithms. *IJCSIS* 2014, 5(3): 328993.
7. McLachlan GJ. Cluster Analysis and Related Techniques in Medical Research. *Stat Methods Med Res* 1992, 1(1): 27-28.
8. Mahr A, Katsahian S, Varet H, Guillevin L, Hagen EC, Höglund, P. *Revisiting the Classification of Clinical Phenotypes of Anti- Neutrophil Cytoplasmic Antibody-Associated Vasculitis: A Cluster Analysis*, 2012.
9. Mooi E, Sarstedt M. *Cluster Analysis: A Concise Guide to Market Research*. Berlin Heidelberg: Springer 2011, 237-84.
10. Norusis M. *SPSS 15.0 Statistical Procedures Companion Cluster Analysis: Upper Saddle River (NJ)*: Prentice Hall 2007, 361-91.
11. Clatworthy J, Buick D, Hankins M, Weinman J, Horne R. The Use and Reporting of Cluster Analysis in Health Psychology: A Review. *British Journal of Health Psychology* 2005, 10(3): 329-58.
12. Anderberg MR. *Cluster Analysis for Applications Academic Press*, New York, 1973, 553-55.
13. Ramazani FT, Rahmani K, Moradi A, Montazeri SA, Bidhendi RY, Darabi F. Comparison of Two Statistical Methods to Determine Normal Range of Androgen Hormones: K- Means Cluster Analysis and Receiver Operating Characteristic Curve, *Journal of Family and Reproductive Health* 2018, 12(2): 98.

14. Megyesiova S, Lieskovska, V. Analysis of The Sustainable Development Indicators in The OECD Countries, *Sustainability* 2018, 10(12): 4
15. Khanmohammadi S, Adibeig N, Shanehbandy S. An improved overlapping k-means clustering method for medical applications. *Expert Systems With Applications* 2017, 67: 12-18.
16. Hartigan JA. *Clustering Algorithms*, NewYork; John Willey&Sons 1975.
17. Dardac N, Boiton IA. A Cluster Analysis Approach for Bank's Risk Profile: The Romanian Evidence, *European Research Studies* 2009, 12(1): 111.
18. Ma Z, Tavares JMR, Renato JN. A Review on the current segmentation algorithms for method images. *In IMACAPP* 2009, 135-40.
19. Johnson RA. *Applied Multivariate Statistical Analysis* 2002.
20. Blei D, Lafferty J. Topic Models. In A. Srivastava and M. Sahami (Eds.), *Text Mining: Classification, Clustering and Applications*. Boca Raton, F. L: Taylor & Francis Group 2009, 71-94.
21. Ada M, Altunay F, Civelek M, Kaplan S, Koç P. Kümeleme Analizi, Başkent Üniversitesi Tıp Fakültesi, tip.baskent.edu.tr/kw/upload/600/dosyalar/cg/sempozyum/.../13.P9.pdf, 22 Ocak 2019.
22. Sharma S. *Applied Multivariate Techniques*, John Wiley and Sons, NewYork, 1996.
23. Selanik M. Türk Tarımının Avrupa Birliği İçindeki Yerinin Kümeleme Analizi ile Belirlenmesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Yüksek Lisans, Ankara: Gazi Üniversitesi 2007.
24. Tatlıdil H. *Çok Değişkenli Analiz*, Akademi Matbaa, Ankara, 1996.
25. Mariott FHC. Practical Problems in a Method of Cluster Analysis, *Biometrics* 1971, 27: 501-14.
26. Bacher J. *Cluster Analysis*, Lecture Notes, Nuremberg 2002.
27. Blei DM. *Hierarchical Clustering*, Princeton University 2008.
28. Norusis MJ. Chapter 16: *Cluster Analysis*. *PASW Statistics 18 Statistical Procedures Companion*, Upper Saddle River, NJ: Prentice Hall 2010, 361-91.
29. Yim O, Ramdeen KT. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data, *The Quantitative Methods for Psychology* 2015, 11(1): 8.
30. Bacher J. *Cluster Analysis*, Lecture Notes. Nuremberg 2002.

31. Gürsakal N. *Sosyal Bilimlerde Araştırma Yöntemleri*, Uludağ Üniversitesi Yayınevi, 2001; 25-26.
32. Florek K, Lukaszewicz J, Perkal J, Steinhaus H, Zubrzchi S. Sur La Liason: Division Des Points D'un Ensemble Fini. *Colloquium Mathematicum* 1951, 282-5.
33. Sneath PHA. The Application of Computers to Taxonomy. *Journal of General Microbiology* 1957, 201-6.
34. Almeida JAS, Barbosa LMS, Pais AACC, Farmasinho SJ. Improving Hierarchical Cluster Analysis: A New Method with Outlier Detection and Automatic Clustering Chemometrics and Intelligent Laboratory Systems, 2007, 87: 208-17.
35. Sokal RR, Michener CD. A Statistical Method for Evaluating Systematic Relationships. *The University of Kansas Scientific Bulletin* 1958, 38: 1409-38.
36. Mazzocchi M. *Statistics for Marketing and Consumer Research*. London 2008.
37. Gower JC. A general coefficient of similarity and some of its properties, *Biometrics* 1971, 27(4): 857-87.
38. Çokluk Ö, Şekercioglu G, Büyüköztürk Ş. *Sosyal Bilimler için çok Değişkenli İstatistik Spss ve Lisrel Uygulamaları*, Pegem Akademi 2010.
39. Rencher AC. *Methods of Multivariate Analysis*, Second Edition, A John Wiley & Sons, Inc. Publications 2002, 451-81.
40. Anderberg MR. *Cluster Analysis for Applications Academic Press* New York 1973, 553-55.
41. Tatlıdil H. *Uygulamalı Çok Değişkenli Analiz*, Ankara: Cem Web Ofset 1996.
42. Servi T. Çok Değişkenli Karma Dağılım Modeline Dayalı Kümeleme Analizi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi, Adana: Çukurova Üniversitesi 2009.
43. Nithya N, Duraiswamy K, Gomathy P. A survey on clustering techniques in medical diagnosis. *IJCST* 2013, 1(2): 17-23.
44. Li X, Zhu F. On Clustering Algorithms for Biological Data. *Engineering* 2013, 5(10): 549.
45. Ding C, He X. K-Means Clustering Via Principal Component Analysis. Computational Research Division, Lawrence Berkeley National Laboratory, Berkeley, <http://ranger.uta.edu/~chqding/papers/KmeansPCA1.pdf> , 13 Şubat 2019.

46. Trpkova M, Tevdovski D. Twostep Cluster Analysis: Segmentation of Largest Companies in Macedonia, Challenges for Analysis of The Economy, the Businesses and Social Progress, file:///C:/Users/PC/Downloads/marija_trpkova__dragan_tevdovski__twostep_cluster_analysis__segmentation_of_largest_companies_in_macedonia.pdf , 18 Şubat 2019.
47. Zhang T, Ramakrishnan R, Livny M. BIRCH: Method for very large databases. *Proceedings of the ACM. Management of Data*, 1996, 103-14.
48. Meila M, Heckerman D. An experimental comparison of several clustering and initialization methods. *Microsoft Research Technical Report MSR-TR*, 1998, 98-106.
49. Chiu T, Fang D, Chen J, Wang Y, Jeris C. A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining*, San Francisco, 2001, 263–8.
50. Corp I. IBM SPSS statistics for windows, version 25.0. Armonk, NY: IBM Corp 2017.
51. Alpar R. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*, Detay Yayıncılık 2017.
52. Kalaycı Ş. *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, Asil Yayın 2009.
53. Banfield JD, Raftery AE. Model based Gaussian and non Gaussian clustering. *Biometrics* 1998, 49: 803-21.

EKLER

EK-1. Özgeçmiş

ÖZGEÇMİŞ

Adı Soyadı: Tuba USLU

Doğum Tarihi: 1990

Öğrenim Durumu: Yüksek Lisans

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	İstatistik	Fırat Üniversitesi	2013
Yüksek Lisans	Biyoistatistik ve Tıp Bilişimi ABD	İnönü Üniversitesi	2019

Yüksek Lisans Tez Başlığı ve Tez Danışman(lar)ı:

Kümeleme Analizi ve Sağlık Alanında Bir Uygulama, Dr. Öğr. Üyesi. Harika Gözde GÖZÜKARA BAĞ

Görevler:

Görev Ünvanı	Görev Yeri	Yıl
Öğr. Görevlisi	Aksaray Üniversitesi	2018-Devam ediyor

EK-2. Etik Kurul Almama Gerekçesi

SAĞLIK BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

13 Nisan 2013 tarih ve 28617 sayı ile T.C. Resmi Gazetede yayınlanan “ Klinik Araştırmalar Hakkında Yönetmelik’in birinci bölümünün 2. maddesinin 1. fıkrası (Bu yönetmelik biyoyararlanım ve biyoeşdeğerlik çalışmaları dahil, ruhsat veya izin alınmış olsa dahi insanlar üzerinde yapılacak olan ilaç, tıbbi ve biyolojik ürünler ile bitkisel ürünlerin klinik araştırmaları, klinik araştırma yerlerini ve bu araştırmaları gerçekleştirecek gerçek veya tüzel kişileri kapsar.) gereğince yüksek lisans öğrencisi Tuba USLU’nun tezinin klinik bir çalışma olmaması, kullanılacak olan verinin web sitesinde yayınlanmış veri tabanından elde edilecek olması sebebiyle Etik Kurul kararı alınmamıştır.